修士論文

Master's thesis

プロジェクタベーストARを用いた人間作業の支援システム

A human-task assisting system using projector based Augmented Reality

中央大学大学院

理工学研究科経営システム工学専攻

山下 雅也

Department of Industrial and Systems Engineering,

Graduate School of Science and Engineering, Chuo University

Masaya Yamashita

2001年3月12日

概 要

次世代ロボットシステムの応用範囲を拡張するために，ヒューマン・ロボットインタラクションのさまざまなレベルにおいて使い易く，親しみやすいインタフェースに発展ことが必要である．

我々は，人間の日常生活を情報強化することができる Augmented Reality(AR) アプローチによるヒューマン・ロボットインタフェースシステム:PARTNER を開発してきた．本論文では，その第 4 のサブシステム，Adaptive Annotation Function(AAF) について述べる．AAF は従来，多く見られる人間からロボットへのタスク教示に用いるシステムとは対照的である，人間の作業をロボットが支援するシステムである．

PARTNER はメッセージ，レンジファインダの投影光などを投影するプロジェクタと状況認識のための CCD および赤外線カメラから構成されるため，対象物の形状変化と人間の動作の両方が認識できる．さらに，プロジェクタからメッセージを投影することで作業者を支援することができる．ポータブル OHP 装置を組み立てる実験を通し，実際の環境で作業者が特殊なデバイスを装着すること無く，ロボットによる作業を支援できる可能性を確認できた．

謝辞

本論文を作成するにあたり，指導担当の中央大学理工学部経営システム工学科坂根茂幸教授に研究の指針や多くの助言，研究者としての姿勢，考え方などの御指導を頂きました．心より感謝致します．そして，生活面において多大な迷惑をかけてしまいながらも，協力を頂いた家族一同にも感謝します.

さらに，さまざまな議論を交わした同研究室の周洪均氏，プログラミングなどに関して多くの情報を下さった佐藤信氏，公私に渡り互いに励まし合った同輩の院生たち，大変貴重な時間を惜しむことなくともにした先輩後輩，いろいろご配慮頂いた技術員の方々にもこの場を借りて厚く御礼申し上げます.

目 次

第1章　序論

1.1　本研究の背景と経緯

高齢化社会の到来，少子化現象など今日の社会では多くの問題を抱えている．そのため，労働者人口が減少する一方で，エネルギーの需要量の増加に応じて人間に代わる高い労働力の開発が必要となり，その代表としてロボットが挙げられる．

今日では，ロボットの機械構造，および，ビジョンの高性能化により，根底となる工学技術の基礎基盤がある程度が確立された．最近のロボット人気の風潮はそのハードウェアの上に簡単な知能をロボットに備えることが可能となったことに起因する．つまり，現在のロボット研究において，従来より上位レベルの人間の思考内容，例えば，抽象概念，予測，計画などの研究が盛んになり，それらの研究結果が世間に受け入れられ始めたのである．

その新しい研究分野の中にヒューマン・ロボットインタフェースが含まれる．従来のロボットは工場における単純な繰り返し作業に用いられていたが，技術が向上するにつれ，人間とロボットの新たな関わり合いを目指す需要が生じた．そして最近になって，ロボットを作業者だけではなく，本格的に人間の代用として利用する試みが工場の生産の場以外にも行なわれようとしている．

われわれは，人間がロボットに簡単に指示を与え，大まかな意図を汲み取り，細かな行動に関しては自分で計画し，指示を遂行するロボットの開発を目指している．この方針の下で現在，ARアプローチによる柔軟性を考慮したヒューマン・ロボットインタフェースシステム:PARTNER (Projector-based AR for Teaching NEw tasks between human and Robot) を開発している．ARとは日常の実空間に対し，電子的に情報強化を行なう技術であり [18]，代表的な方法として，Head Mounted Display(HMD)[5] を用いる方法が挙げられる．しかし，HMD は装着型デバイスであるため，使う際には身に付けねばならず，実用性に乏しい．PARTNER のプロトタイプシステムは情報表示のためのプロジェクションシステムと，環境と人間のアクションを認識するための実時間トラッキングビジョンシステムから構成される．プロジェクタを用いた AR の例は DigitalDesk[16] などがあり，装着器具を用いずに柔軟なインタラクションを実空間に実現が可能であることを示した．しかし，DigitalDesk で開発されたアプリケーションはオフィス用途に留まり，実環境に用いる人間とロボットのインタフェースに，利用された例は少ない．

1.2　本論文の目的

PARTNER のインタラクティブサブシステムとして現在までに，Virtual Operational Panel, Interractive Image Panel[14]，Interactive Hand Pointer[21][12] の３種類を開発してきた．これらはロボットタスクの教示を目的としているのに対し，本論文では，第４の機能として，人間の作業を支援するための Adaptive Annotation Function(AAF) について述べる．

具体的には，PARTNER の構成要素であるプロジェクタと監視カメラから構成され，これらの要素を用いて，人間の作業を AR により情報強化を行なう．さらに，作業による状況変化を認識し，その都度，その状況に応じた支援を行なうことを目的としている．

1.3　本論文の構成

本論文では，研究の目的を明確にするため，第２章では従来のヒューマンインタフェースの従来のあり方について述べ，近年，開発が盛んな AR アプローチによるヒューマン・コンピュータインタフェースを紹介し，それに対する考察を行なう．第３章では，我々が提案しているヒューマン・ロボットインタフェースシステム: PARTNER について述べる．ここでは，ロボットへの作業教示を目的とし，これまでに開発を行なった VOP，IIP，IHP の三つのインタラクションサブシステムについて簡単に紹介する．

第４章では今回開発した PARTNTER の第 4 のインタラクションサブシステム Adaptive Annotation Function について述べる．この章では，従来の人間の作業支援を行なうシステムとの相違点を明確にさせ，その支援方法を具体的に示す．この AAF による作業教示をポータブル OHP を組み立てる作業に適用し，その有効性を確認し，考察をまとめる．その上で，第５章で本研究で得られた成果をまとめ，今後の課題を述べる．

第2章　Augmented Realityを用いたヒューマンインタフェース

本章では，従来の人間とコンピュータ間のインタフェースについて述べ，新しいアプローチ方法として注目されている Augmented Reality を用いたインタラクションシステムについて述べる．

2.1　AR 以前のヒューマンインタフェース

1990 年代後半より，コンピュータの使用人口が増大し始めた．これは，一般の人々に対して，扱いやすいように設計されるようになったからであり，それにともない，インタラクション機能の向上が著しく発達した．今日において主流となっているコンピュータとのインタラクション方法は，主にコマンドユーザインタフェース (CUI) とグラフィックユーザインタフェース (GUI) である．以下に，これらの特徴を挙げる．

CUI

1. 熟練者以外は操作がわかりにくく，トラブル回避が難しい
2. コマンド入力ミスが起こりやすく，エラーをしやすい
3. 操作の習得に時間がかかる

GUI

1. グラフィックによるディスプレイが可能であり，２次元，３次元のオブジェクトを理解させやすい
2. CUI に比べ，操作方法が容易になったが，何度もクリックが必要になる
3. コンピュータディスプレイのなかでポインタで狙いを絞る必要がある

コンピュータの使用者はワープロ作成，ウェブブラウジング，電子メール，プレゼンテーション，ソフトウェア開発とさまざまな用途に用いるようになり，GUI と CUI を共存させる X-window などのインタフェースシステムが利用できるようになった．しかし，依然として高齢の入門者にはコンピュータの評価が高くない．この原因は，人間がコンピュータの接触方法に合わせているため，つまり，不自然なインタラクション方法だからである．

これに対し，さまざまなコンピュータとのインタラクション方法が開発されてきた．

- タッチパネル画面からの入力システム

- 音声認識システム [2]

- 文字認識システム [31]

これらは表示機能のみのコンピュータディスプレイを介した従来のインタラクション方法とは異なり，オペレータに対して新たな技術を用いて人間同士のコミュニケーションに近い入力方法を実現した．しかし，人間同士のような気軽なコミュニケーションにはまだ遠い．コンピュータがさらに人間を察し，状況に応じて適当な出力を行なう機能が望まれている．この問題の解決をはかり，次世代インタラクションシステムとして AR を利用したものが近年，研究されている．

2.2 Augmented Reality

ヒューマン・コンピュータインタフェースの分野では，実環境のマルチ-モーダル情報を統合することによって，人間の日常生活に用いる知覚への影響を強化し，現実感を与える Augmented Reality(AR) システム [11] あるいは Mixed Reality システム [17] が注目を集めている．これらのアプローチは人間が日常用いるコミュニケーション手段に近いインタフェースシステム構築への指針を示し，ヒューマン・ロボットインタフェースシステムにおいても同様のことが言える．

ここで挙げた AR とは仮想現実感 (Virtual Reality:VR) から派生した．AR は３次元コンピュータグラフィックスなどの仮想環境を実空間に融合する技術である．この技術はコンピュータの画面の前に座ってキーボードやマウスで入出力する従来の GUI とは異なる，実世界指向のインタフェースの可能性を見いだした．

以下で，ハードウェア別の利用方法を説明し，特徴を述べる．

HMD による AR

AR に用いるもっとも有名なデバイスとして Head Mounted Display(HMD) がある．HMD は近年，ウェアラブルコンピュータとして次世代型コンピュータの代表となっている．本来，VR の生成する CG を眼鏡のグラスに投影して，臨場感をオペレータに誘発することを目的に作られた．この HMD のグラスを透過ディスプレイにして外界を眺める．この時，ディスプレイに情報を投影することで AR が実現できる（図 2.1）．

HMD による AR システムはさまざまな利用がされている．医療応用の分野では CT スキャナなどから得られた妊婦の胎内のデータをを see-through Head Mounted Display をとおして覗くと超音波センサから得られたデータをもとに，Display 上に超音波像をスーパーインポーズして胎児が動いたり，蹴ったりしている様子を同時に提示する技術を開発した [3]．また，人間の作業支援システムとしては，Feiner らが作業対象物のプリンタに備えた磁気センサから位置情報を得て，作業方法を三次元モデルや矢印の表示により支援する KARMA[5] を開発した．また，ボーイング社はワイヤーの結線作業において利用し，結線すべきワイヤーのみをディスプレイにインポーズして作業の効率成果を上げている [1].

see-through Head Mounted Display を用いた AR システムは視点情報を基に表示情報を生成する必要があるため，HMD 装置の追跡処理が必要となる．また，装着型デバイスであるため，不自然な状況において作用することになる．

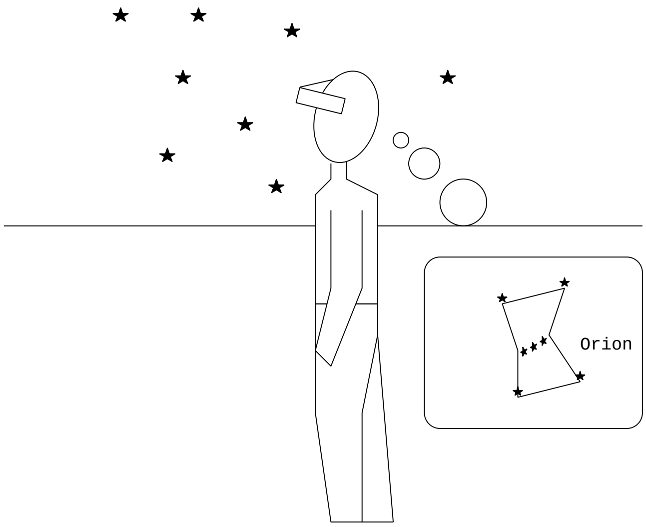


図 2.1: AR system using Head Mounted Display

ハーフミラーによる AR

ハーフミラーを用いて，AR を実現するシステムが伴らによって開発された [25]．これは作業台上の電子機材の作業箇所を示し，たとえば，つまみのまわす向きを表示するなどの作業方法や計測器のプローブ２本をつないだ時の抵抗値の情報をディスプレイに表示するバーチャル電子計測器である．この方式は透かして見える実空間とミラーに反射して見える仮想世界を重ね合わせることにより実空間強化を試みる（図 2.2）．

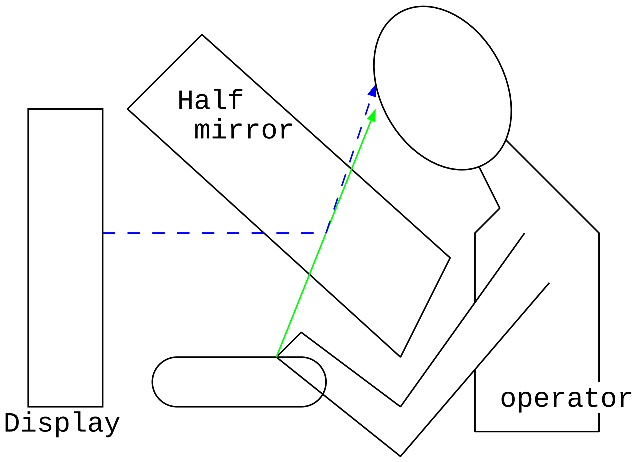


図 2.2: AR system using half-mirror

このシステムは固定されたハーフミラーを通して作業を行ない，ミラーの下に手を入れ，ミラーを決まった角度からのぞき込む必要がある．そのため，作業を行なう場所を限定され，実用性が高いとはいえない．

プロジェクタによる AR

Wellner らが構成した DigitalDesk[16] が先駆けとなっている．机の上にさまざまな記号やパターンを投影し，それに対する人間の手の動作認識というサイクルを繰り返して，柔軟なインタラクションが可能となることを示した．具体的には机の上にコンピュータの画面が投影され，ユーザは机の上に置かれた実際の紙やその上に書かれた文字や絵とともに，コンピュータの映像を見ることができる．また，逆に机の上にコンピュータの表示する文字や絵を投影することが可能である．

Sato らは電子基盤の製作作業の支援システムとして指さした対象物を認識させたり，どの対象物に対して手を伸ばしているかを認識させるシステムを実現した [15]．また，小林らの Enhanced desk[19] では，赤外線カメラを用いて実時間指先認識を実現し，マウス操作のクリックなどを指を使って代用するシステムを実現した．

その他に，モニタディスプレイによる AR は，HMD 遠隔操縦をする際の作業で，松井はロボットや環境の 3 次元幾何モデルを実画像の上に重畳するマルチメディアディスプレイ: MMD[23] を開発した．これはシャッタレンズを通して，ディスプレイを見ることで立体視を実現し，マニピュレータによるバルブ弁の交換作業の遠隔操作を身近に見えるようなテレオペレーションの監視を行なった．同様に Rastogi らが開発した ARGOS[4] も HMD を通してモニタディスプレイを見る立体視映像から，遠隔操縦でマニピュレータを操作して積木作業を行なった．

2.3 AR のヒューマン・ロボットインタフェースへの適用

以上で述べた AR システムは対象物に情報を重ね合わせるように投影し，目をそらすことなく，実際の環境で支援を行なうことができる．これにより，従来のモニタディスプレイを見つめるインタラクションから離れ，人間の自然なジェスチャによる意志伝達が可能である．しかし，装着型のデバイスを用いる AR については 3 次元モデルを重ね合わせるために人間の頭部位置を計測し，視点を推定する必要がある．HMD やハーフミラーによる AR ではそのような理由から，頭部を固定したり，追跡したりする必要がある．作業を行なう際も，機材を設置するために，人間に無理な作業を強いることがあり得る．HMD を用いる場合には，機材の重さ，身に付ける不自然さから違和感を感じ，ハーフミラーを用いる AR では，ミラーを通して見る必要があるため，手をミラーの下に入れて作業をする窮屈な姿勢を強いられる．

それに対し，プロジェクタベースト AR では外部からの情報投影となるため，以下の特徴がある．

- プロジェクタにより直接投影するため，傷つけることもなく，オペレータに対象オブジェクトを理解させやすい．
- 装着型デバイスでは環境を多人数に影響させる場合，全員に違う映像を提供しなければならない．そのため，計算機に多量の計算コストがかかってしまうが，本システムではオペレータの都合に左右されず，システムの簡略化が計れる．また，作業者が頻繁に代わる状況でも対応できる．
- 特別なセンサ類を体に装着したり，持ち歩く必要が無く，自然なインタラクションが実現できる．

- プロジェクタとそれを監視するカメラがあれば十分なインタラクションが可能である．

このシステムはプロジェクタの投影範囲内でなければ効果がない．また，日光の強い屋外では実用が難しいが，自律機能を持てば，室内のインテリジェント照明として十分に期待できる．

従来のロボット操作は専用のコントローラでその場で直接，監視しながら操作するか，あるいは，コンピュータに動きを入力してその動きをフィードバックさせるものがほとんどであった．近年，ロボットを専門家以外のユーザーが手軽に扱う需要が増えたため，AR アプローチによるロボット操作もこうした流れを積極的に採り入れる必要がある．そこで，プロジェクタベースト AR は実際の空間に反映されるため，実世界指向の新しいヒューマン・ロボットインタフェースとして期待できる．

第3章　PARTNER

3.1　ロボット作業の教示システム

プロジェクタを用いるARには，HMD等の特別な機器を身体に装着することが不要で，また，複数の人間が融合環境を共有できるという利点がある．そこで我々は，新しいヒューマン・ロボットインタフェースシステムの手法として，プロジェクタベーストARを用いたインタラクションシステム: PARTNER(Projector-based AR for Teaching NEw tasks between human and Robot)を提案している．PARTNERは人間-ロボット間の４つのインタラクションサブシステムから構成される（図3.1）．

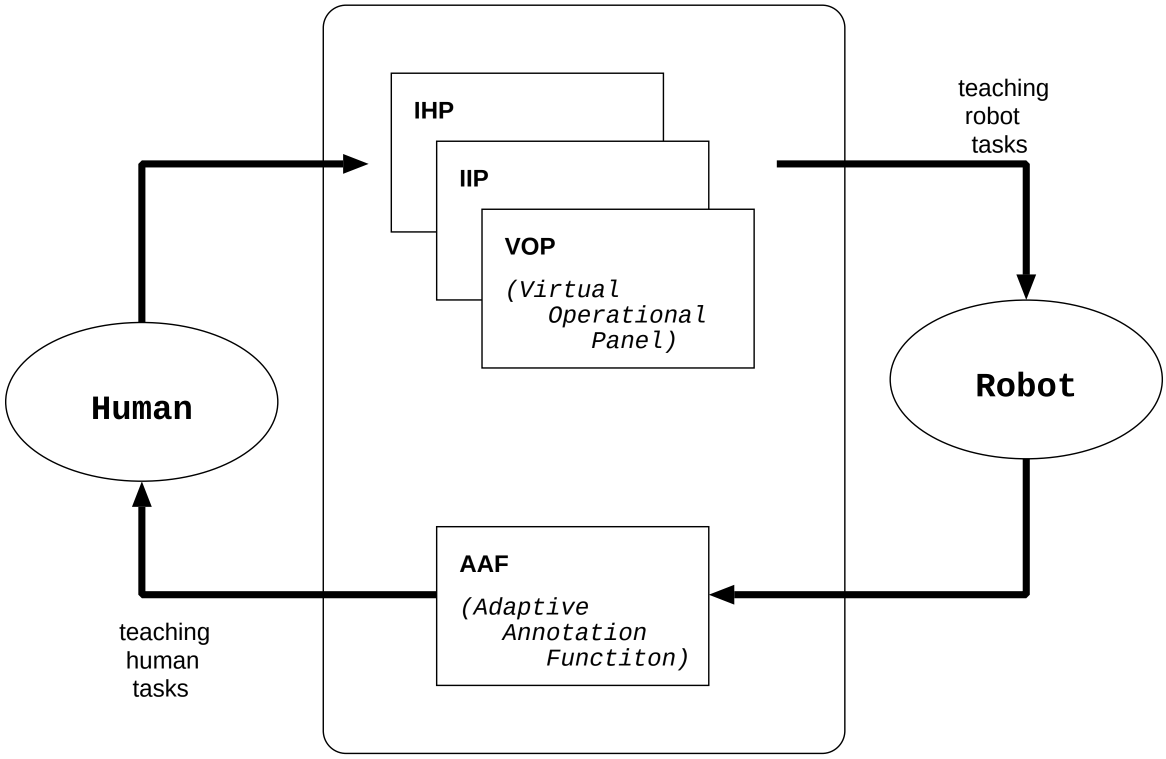


図3.1: Adaptive Annotation Function and other modules in PARTNER

PARTNERのこれまでに構築したサブシステム(VOP, IIP, IHP)は人間からロボットへの作業教示を実現するツールとして開発した．それと対照的にAAFはロボットから人間に作業を支援するシステムである．本章ではVOP，IIP，IHPについて述べる．

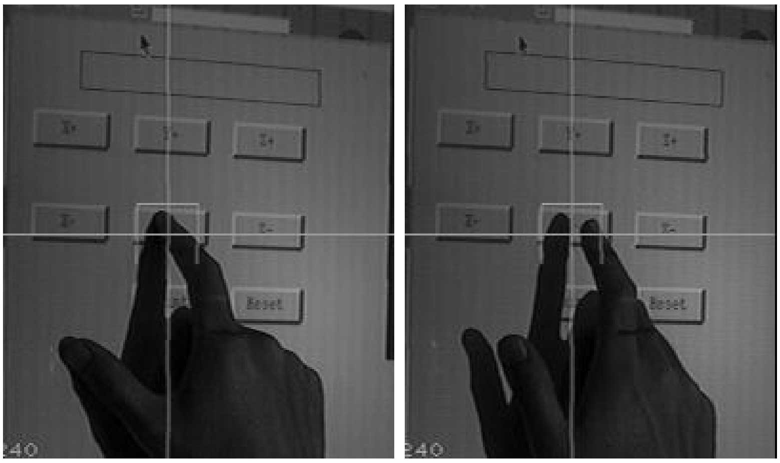


図 3.2: Virtual Operational Panel

3.1.1 Virtual Operational Panel

寺島らの開発した Vertual Operation Panel[14] では，作業台に仮想操作パネルの投影を行ない（図 3.2），オペレータの指先をトラッキングビジョン装置によりの追跡し，パネルに対して接触したかを判定してロボットに作業教示を実行する．このとき，投影されるパネルは銀行の ATM などのタッチスクリーンデバイスに見られるような button アイテムだけでなく，slider, joystick，infomation window も用意している．さらに，パネルはアフィン変換 [22] を用いてその状況に応じた図形に変換して投影することも出来る．

3.1.2 Interactive Image Panel

VOP を発展させた IIP は，机の上に作業環境の実画像を直接指し示すことでロボットとのインタラクションを行なう [14]．これにより，以下のような作業が可能である．

1. 作業者が指示したい対象物を指し示すか，その画像のある領域を指定する（図 3.3）
2. システムがトラッキングビジョン装置によりその対象物を同定し，結果を画像上に表示する
3. 画像上に提示された情報を作業者が確認する
4. 指により対象物を移動したい場所まで指示する
5. 指を離した位置が例えば机であればロボットは対象物をその場所に移動させる

本機能は，実際の画像に指をさして，ロボットを操作するため，視線を移すことなく指示を行なうことが出来る．

3.1.3 Interactive Hand Pointer

前述の VOP と IIP は 2 次元平面に拘束された指示動作に用いるのに対し，佐藤らの開発した IHP[21] は作業空間における 3 次元的な指示動作が可能である．2 台のカメラの入力画像から

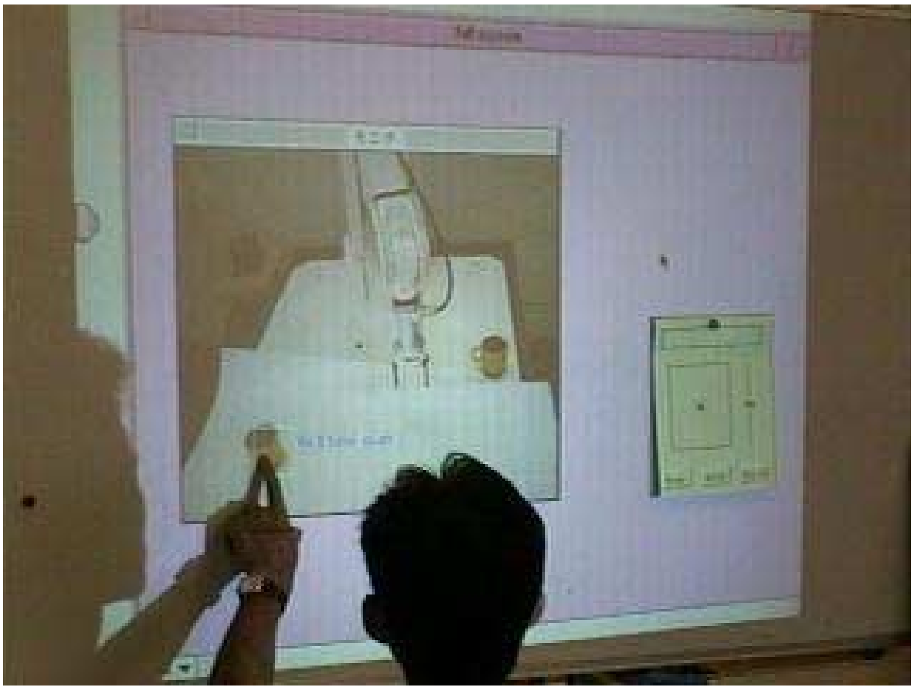


図 3.3: Interactive Image Panel

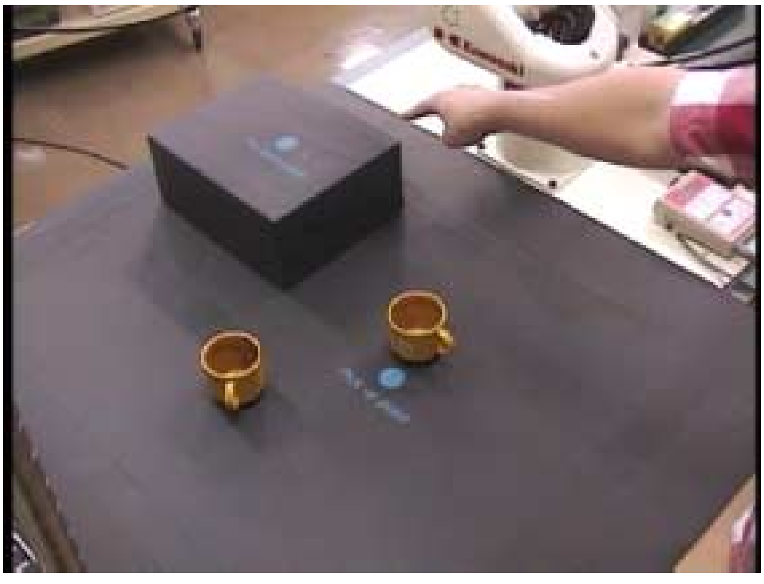


図 3.4: Interactive Hand Pointer

指先，および，指の付け根を探索し，カメラのそれぞれの画像内での指先方向を検出する．その検出データを用いて，プロジェクタから対象にマークを投影する．このIHPはプロジェクタと2台のカメラから高さ推定機能を備えているために作業台の上のみでなく，対象物の上にもマークの投影が出来る．これにより，カップの移動などの動作をロボットに指示することが出来る（図3.4）．

第4章　Adaptive Annotation Function

本章は，VOP，IIP，IHP に続く第 4 のサブシステム：Adaptive Annotation Function(AAF) について述べる．

4.1　AAF の特徴

人間が新しい機器を使う際に，通常は取扱説明書を読まなければならない．この場合，説明書と実物を自分で対応付ける工夫，たとえば，説明書の置き場を目の届きやすい場所に設定することを考えたり，実物を片手で掴んで場所を忘れないようにしながら他方の手で説明書のページをめくる作業が必要である．それに対して，実際の作業の時に実物に直接，説明を記述できれば作業効率が良くなることが予想できる．

AAF では，PARTNER の構成要素を用い，プロジェクタを用いて作業を実物に投影することで，人間作業の支援システムを構築した．組み立て作業のための adaptive annotation を含む状態遷移は図 4.1 を使用する．各ガイドオペレーション GO_k は初期アノテーション A_k によって始め，タスク状態のチェック（監視）を行なう．もし期待する状態に至れば，その時に状態 S_k に移る．さもなければ（状態 S'_k，望ましい状態へ導き，ユーザーの誤操作を警告するために第 2 のアノテーション A'_k が生成される．

2.2 節で挙げた AR を用いた作業支援との相違点を以下に記す．

1. PARTNER はプロジェクタを使用するため，モニタディスプレイを見るために see-through HMD のようなヘッドギアを装着する必要が無い．システムは複数のユーザーにアノテーションを示すことに適する．

2. システムは環境の幾何学的変化の調査のためのレンジファインダ機能を使用する．PARTNER のレンジファインダ機能を利用できるため，専用のハードウェアを付加する必要が無い．また，環境認識ではプロジェクタ投影によりラベリングするため，対象物にセンサをとりつけることもない．

3. ユーザーの知識の違いにより，誤操作をする可能性がある．人の行動と，物体変化を監視することでアノテーションシステムが作業の進行状況に応じて支援する．たとえば，操作の仕方に戸惑う初心者に対して，システムはさらにアノテーションを生成する．

4. 赤外線画像は背景のちらつき，色変化，テクスチャに対して影響を受けない．そのため，システムは人の手の行動を信頼性良く調査し，監視することが可能である．

5. グラフィック画像を対象物に直接投影できるため，文字列のみを表示するより直観的でわかりやすい表現ができる．

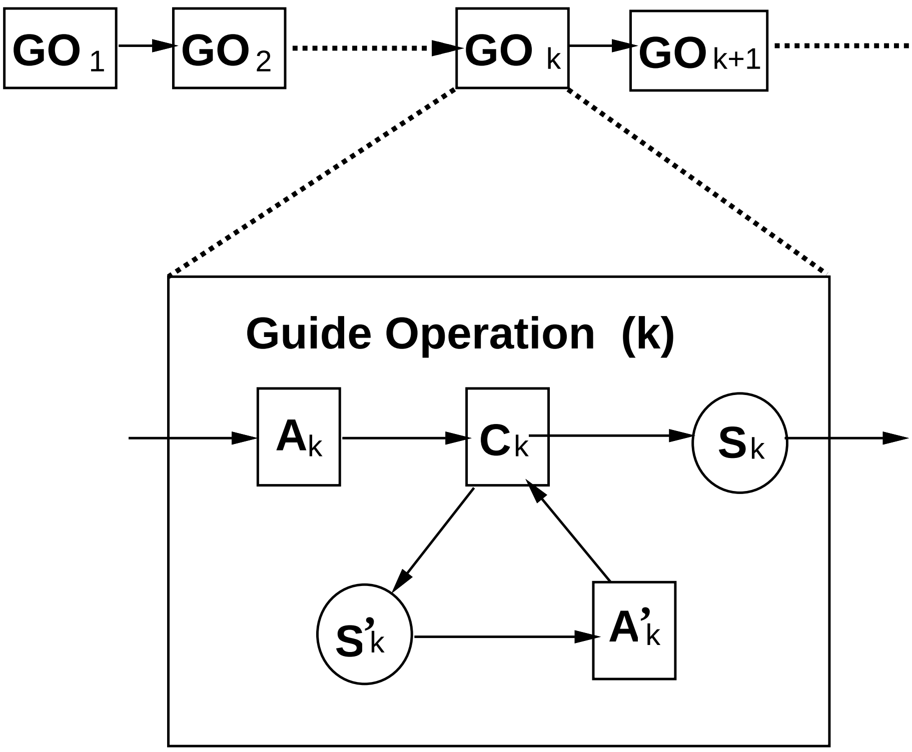


図 4.1: State transition diagram of guiding operations and adaptive annotations(GO$_k$:guide operation, A$_k$:annotation, C$_k$:check of status, S$_k$:status)

6. センシングの際には，環境の空間座標のデータを基に算出する．そのため，より実践的な作業支援が行なうことができる．

7. 図 4.1 に示すような長い一連の作業にも支援が可能である．

グラフィック表示による人間作業の支援としては Feiner らが HMD を用いる支援システム KARMA[5] を開発した．これは対象となるパーツの位置を検出し，その状態に対してアノテーションを行ない，上記の 5 において共通している．しかし，対象物体に取り付けた磁気センサからの情報から位置情報に関してのみに反応し，人の動作に関しては関与していない．さらに，支援する場面も KARMA では給紙の作業にのみ対応していて，図 4.1 のような連続する作業には対応できない．このため，上記の 1 の他に，3 と 7 の点で大きく異なる．

さらに，プロジェクタベースト AR である [15][20] と比べ，本システムはレンジファインダ機能を持ち，ユーザーの作業と物体変化の両方を監視している点で異なる．

4.1.1 物体の状態認識

AAF では，作業の進行具合に応じて支援するため，物体の形状変化など，環境を監視する機能が必要となる．本システムの場合，プロジェクタを用いて情報投影を行なうので，その機能を拡張すれば各種の光パターンの投影に基づくアクティブレンジファインダの構成が可能になる．例えば,（１）点図形の投影によるスポット光レンジファインダ,（２）多数点の投影によるランダムドットステレオ [24],（３）線図形の投影によるスリット光投影レンジファインダ [26] などのレンジファインダが構成可能である．従って，オペレータは多様なレンジファインダ方

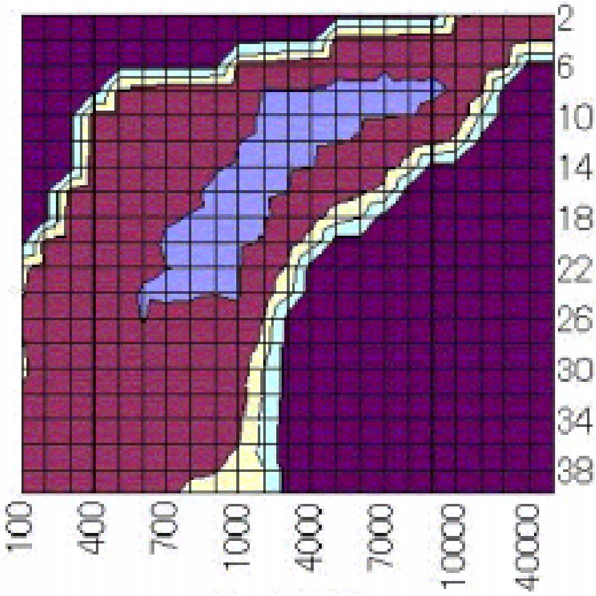


図 4.2: The relationship between the number of dots and diameters

式を用いて，対象物の形状や状況に応じて選択するシステム化（仮想操作パネルの利用），指さしポインタ機能と結合した部分的な３次元情報の入力システム化への拡張が可能となる．

以下ではランダムドットステレオレンジファインダ，スリット光レンジファインダについて述べる．

(1) ランダムドットステレオレンジファインダ

ステレオ視における重要な課題は左右両眼の対応点探索問題であるが，ランダムドットパターンの投影により誤対応を減らすことができる（図 4.3）．対象物の測定面が一様で画像特徴のない面でも，その面に模様をつけることに相当するため，エッジ画像に基づくステレオ法と比較して多くの対応点の情報が得られる利点がある．また，相関演算を用いるレンジファインダ方式なので，トラッキングビジョン装置 [30] との相性が良い．

ランダムドットを投影したシーンに対して，ステレオ視により 3 次元計測を行なう．右カメラ画像 (640×480pixels) において，32×32 pixels を 1 ブロックとする領域を画面縦横に 8pixels ずつ移動させたものに区分けする．横 69× 縦 49 の合計 3381 ブロックを各参照画像とし，ブロックサイズ 8×8，間引き 3 で左カメラ画像との相関演算を行ない，対応する画像上の位置を求める．両カメラの対応点の位置と，予めキャリブレーションにより求めたステレオの変換行列に基づいて 3 次元位置を計算する（付録 A）．相関の精度はランダムドットの径の大きさと分布密度に依存するので，誤対応の発生頻度が少ないドットの径と密度の組合せを調べ（図 4.2），ドットの径を 10pixels(実空間で 12mm 相当)，ドット数を 2500 個とした．図 4.2 は，縦軸＝ドットの径，横軸＝ドットの数，となっている．投影したドットパターンの実空間での大きさは約 920mm ×700mm の領域である．

図 4.4 は図 4.3 をこの方法で求めた作業環境の 3 次元データの例を示している．垂直方向の軸は Z 座標値である．水平方向の 2 軸は X,Y 座標値ではなく，測定データを格納する配列 (69×49) のインデックスの値そのままで示している．全画面にわたる相関演算と対応点の 3 次元座標の計算に要した時間は約 1.33 sec であった．



図 4.3: Project the random-dots

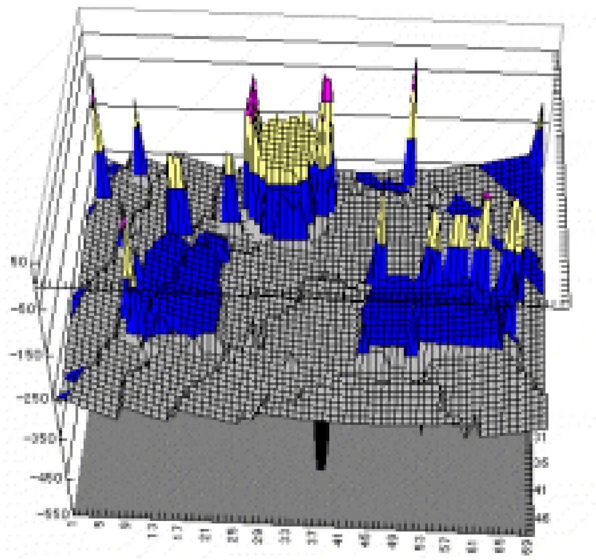


図 4.4: An example of 3D data extracted by the range finder

図 4.5 は上記のシステムにより抽出して物体上にラベル A,B,C を実空間に投影した場面を示している．具体的な処理は，ランダムドットレンジファインダシステムにより 3 次元位置を計測したデータ配列 (69×49) の高さ方向-250mm（作業台の高さ）以上 200mm 未満の計測値を物体の上面と仮定し，2 値化する．この出力データに Edge Preserving Smoothing 処理 [7] を 2 回行ない，その抽出領域（連結成分）にラベルをつける．その上で領域の重心となる配列要素の 3 次元座標値にプロジェクタから物体識別のラベルを投影している．また，連結成分毎に検出するため，対象物体がいくつあり，どこにどちらの方向に向いているのかが検出できる．たとえば，似たような細長い二つの物体を認識し，その片方をマニピュレータにより掴み上げ，もう片方に重なるように向きを合わせて置く，などのロボットへの作業教示が可能である [27].

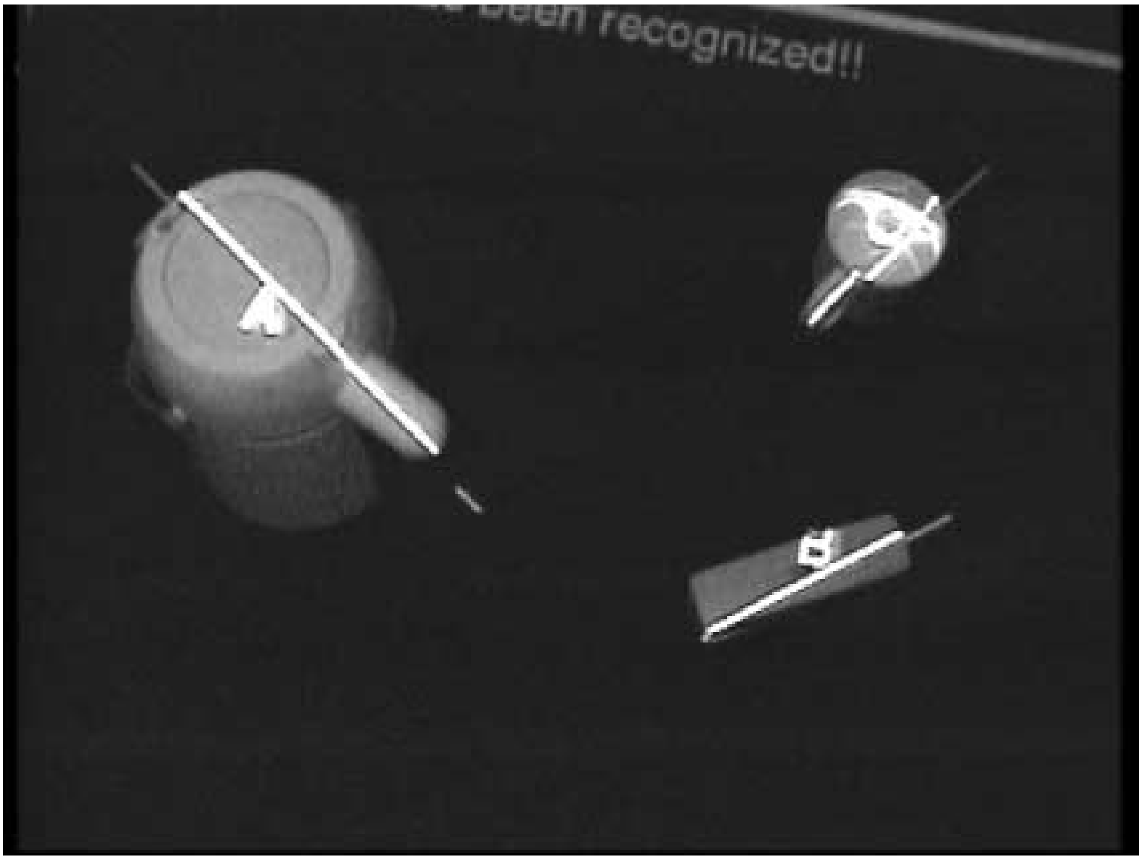


図 4.5: Projection of principal axis to the objects

(2) スリット光レンジファインダ

従来のスリット光レンジファインダは例えば，円筒レンズを用いてレーザー光をスリット状に変形し，光の通った時間と画像面上での座標値から絶対座標を抽出する方法を用いていた [28][26]．これでは，スリット像の方向などを指定できない．しかし，近年のプロジェクション機器の性能の向上により，環境に対する複雑なイメージの比較的高輝度での投影が可能となった．このプロジェクタを用いることにより，従来の投影パターンの異なるレンジファインダ方式が可能となる．

本システムでは，スリット光の投影（従って計測）範囲の局所化と，スリット光の投影角度が可変なレンジファインダ方式の構成を行なった（図 4.6）．

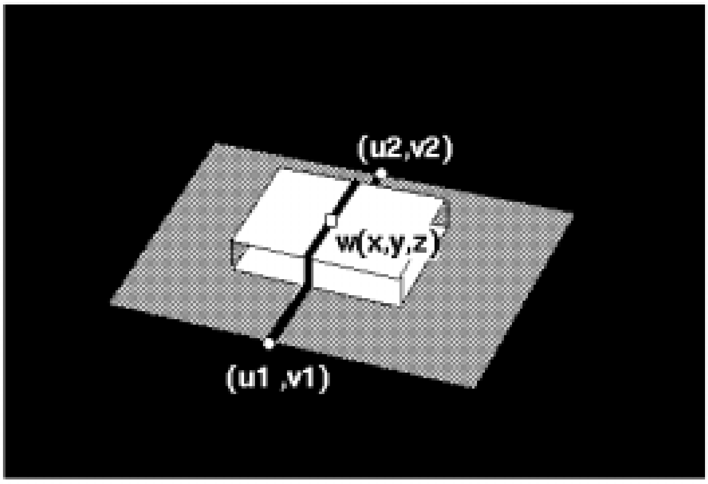


図 4.6: Camera image of using light-striped rangefinder

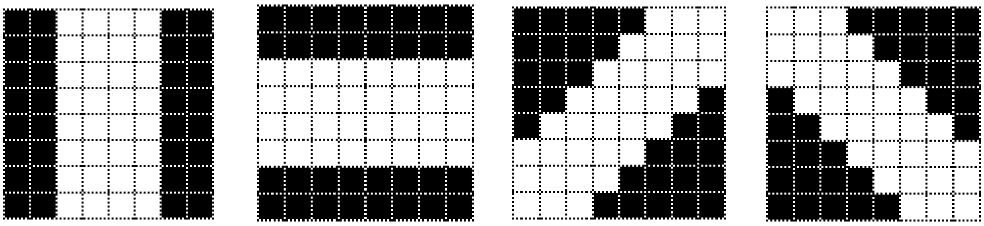
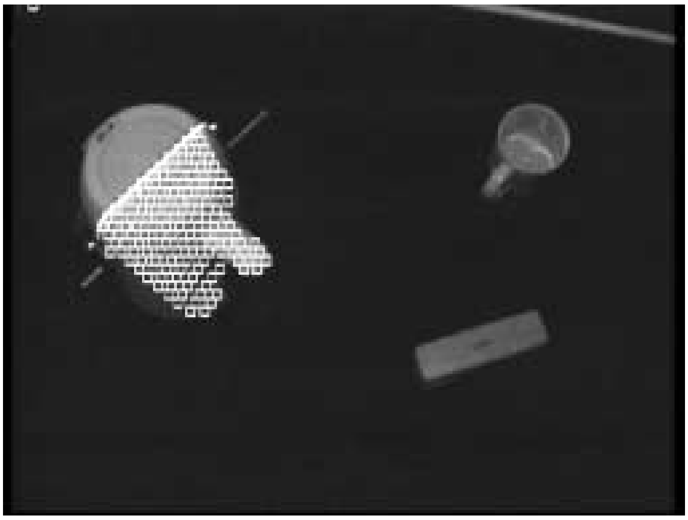


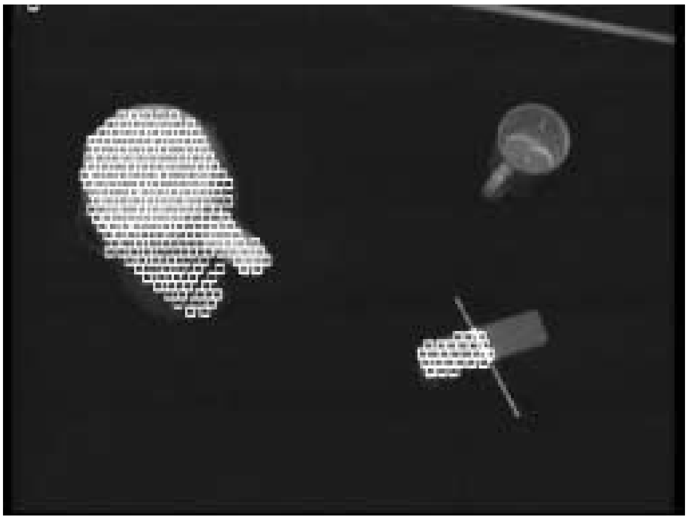
図 4.7: The template images of light-striped rangefinder

上で求めた連結成分の主軸方向に向いた外接長方形の頂点の絶対座標値 2 点をプロジェクタ画像面上に変換する．この 2 点を結ぶ線分上をスリット像の中点が移動するように，さらに，画像面上 2 点間の水平，垂直方向のベクトル成分のうち大きい値をスリット線分の長さとして設定する．これにより投影されたカメラ画像面上のスリット像の位置を検出し，絶対座標値を算出する．この時の座標値計算のアルゴリズムは付録 B にて記す．なお，スリット像の位置の検出方法は事前に描画した 8×8 pixels のスリット像 (図 4.7 の計 4 枚) を参照画像とし，横方向について相関演算をし，最適となった部位を検出する．これを 8pixel ずつ上下にずらして全画面を探索した．

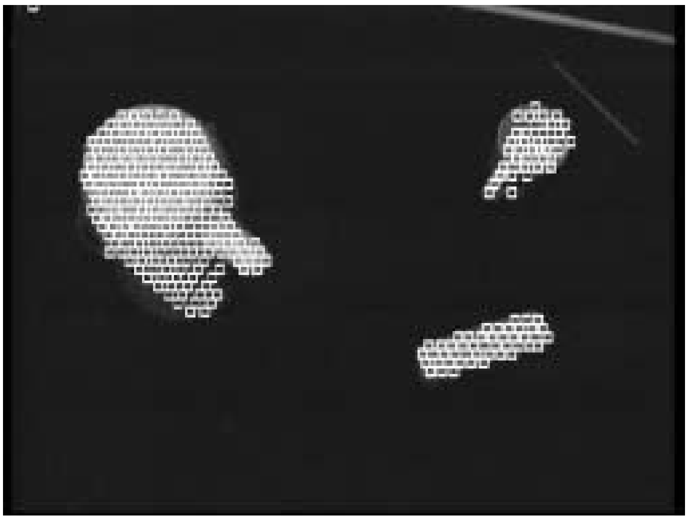
図 4.8 はスリットを画像内で探索し，特定したカメラ画像上の位置をマス目状に描画している様子である．また，ランダムドットレンジファインダでは上面が平面状でない物体を検出することが困難であったが，(c) においてコップの縁部分も検出されていることが示されている．



(*a*)



(*b*)



(*c*)

図 4.8: Camera image of using light-striped range finder

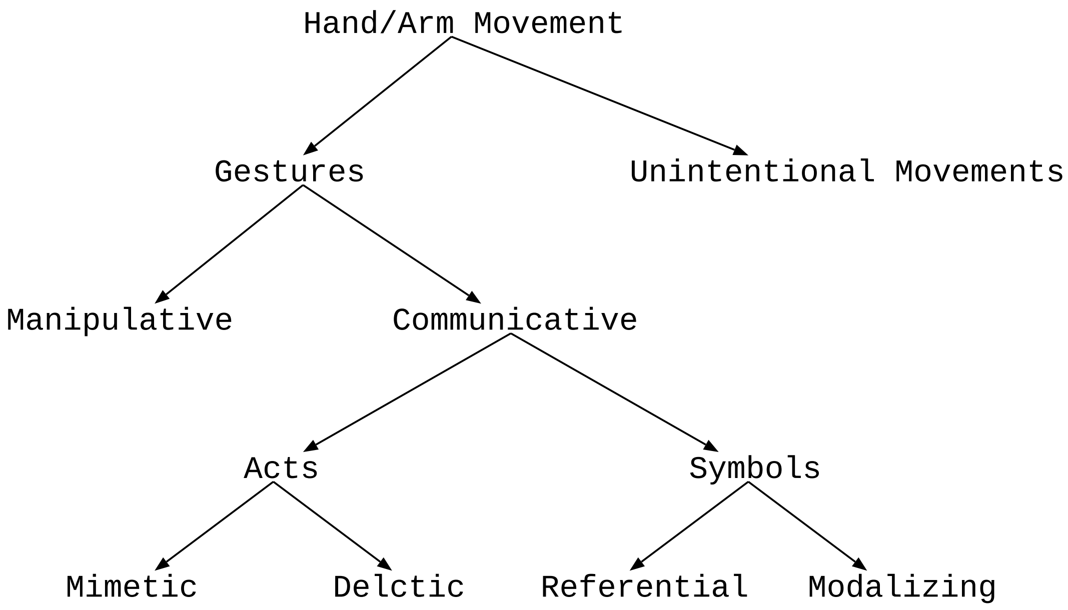


図 4.9: A taxonomy of hand gestures[9][10] for HCI. Meaningful gestures are differentiated from unintentional movements.

4.1.2 人の動作の認識

人の動作認識は，おもにジェスチャ認識の研究が盛んである．たとえば，手話認識 [13] テレビのチャンネルなどをコントロールする [32]，といった意志伝達を目的としている．ジェスチャは Quek によって図 4.9 の様に分類されている [10][9]．まず，意味をなさない動作である Unintentional Movements に対し，Gesture は物体操作である Manipulative と伝達目的に使用される Communicative に分けられる．さらに，真似する動きである Mimetic，強調動作である Delctic に分類される手の動き自身に解釈がある Act と，繰り返し動作 Referenctial や言葉に付随する Modalizing など，静的な姿勢で動かし，意図を表す Symbols に分かれる．

現在，多く行なわれているジェスチャ認識は，Symbols にあたる静的な手の動作の認識研究である．とくに，Referential は繰り返し動作を行ない，時間軸に沿って，一定の動きをするため，隠れマルコフモデル（HMM），動的計画法（DP）によって，リアルタイム処理で認識することができる．これは，観測された手の動きがある一定のルールで動いていることが前提であり，事前に用意した動きのテンプレートの中でどれに近い動きなのかを検出するパターン認識の問題としてとらえられる．

本研究では，人間が物体を操作する様子の認識が必要であるため，Manipulative を検出することが目的となる．池内らは人間が行なう作業を HMM に基づく解析を用いて学習させ，ロボットに同じ作業を行なわせた．[6]．しかし，この場合はデータグローブを装着し，指の関節角度を入力データとして扱うため，自然なインタラクションとは言えない．本研究では，画像情報による手の動作認識を行なう方針だが，これは見る場所によっては見えない場合も起こり得る．特に，”Manipulative” は物体操作を行なうため，物体により隠蔽される可能性もある．そのため，指先が隠れているにも関わらず，推測できるシステムを構築する必要である．

CCD カメラなどのビジョンによる動作認識としては，Pavlovic が Dynamic Bayesian Networks(DBN) を枠組とする modified mixture of DBN モデルにより，手をひらを上下，左右に動かし，“Move up”，”Move left” の Referential の識別などを行なった [8]．この DBN を枠組

とする動作認識はその動作の１フレーム前の時系列データと，その時刻に関わりのある要素から依存していることを前提とし，幅広い認識の可能性を期待できる．

本論文では，指先を追跡し，手がどのあたりにあるのかを考慮して，ロックに近いかを検出している．この場合，手がカメラからつねに一定の状態でなければ検出が難しく，手をひるがえしたり，手を握ることによって指先が検出できなくなる．そのため，不自然な状態での検出に留まっている．今後は，画像データ入力に基づいて，DBN を枠組とする”Manipulative” な動作の認識を行なう必要がある．

4.2 システムの構成デバイス

プロトタイプは情報ディスプレイのためのプロジェクションサブシステムと環境および，オペレータである人間のアクションを認識する実時間トラッキングビジョンサブシステムから構成される (図 4.10).

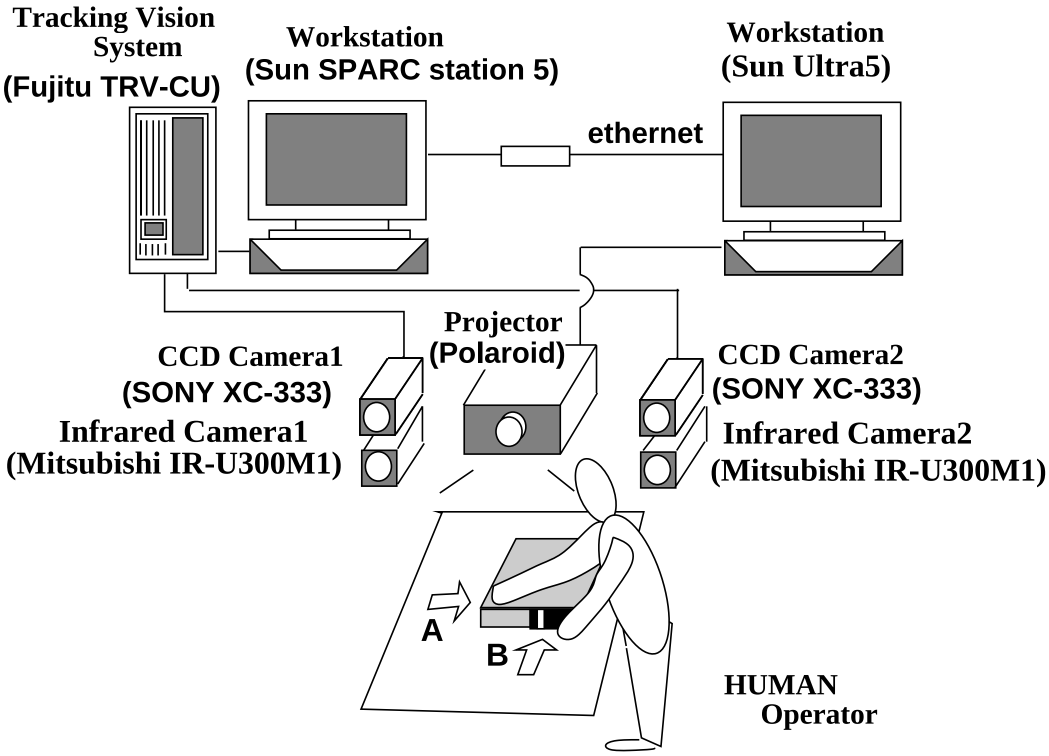


図 4.10: The prototype system of PARTNER

PARTNER では，３台のワークステーションを各デバイスのコントローラとして用い，LAN により各デバイスから得た情報のやりとりをしている．以下では PARTNER の構成要素であるトラッキングビジョン装置，プロジェクタなどを紹介する（図 4.11）.

CCD カメラ

物体検出，距離画像計測などに使用し，トラッキングビジョン装置に入力される．カメラはソニーの”XC-333”，二台を使用している．このカメラは位相チェック機能を持っているため，複数台カメラによるステレオ視に適している．



図 4.11: CCD cameras, infrared cameras and projector

赤外線カメラ

人間の動作監視に使用し，トラッキングビジョン装置に入力される．CCD カメラによる動作監視では，外界の背景色が肌の色に近いと検出が難しい．このカメラによる赤外線画像では体熱を発する人体と背景がのコントラストが明確になり，動作監視に有効である．また，本システムではプロジェクタを使用しているため，環境の光量変化が激しく，CCD カメラでは例えば，指先をテンプレートとした場合，プロジェクタの照明によって追跡を外してしまうおそれがある．赤外線カメラはそれを補うための効果も持つ．使用したカメラは三菱電機製の “IR-U300M1” を使用している．検出波長帯は 8-12μm であり，非冷却型固体撮像素子を用いて赤外光を検出する．

プロジェクタ

Euslisp の Xwindow インタフェース機能を利用し，Sun Ultra5 上の X サーバーに描画する．プロジェクタによりその出力画面を投影することによりイメージ投影，実験台上のオペレータに対するメッセージ送信などが行なえる．事前にキャリブレーションを行なっているため，プログラム上で作った幾何モデルを実物体に重ね合わせることも可能である．使用したのは Polaroid 製”COLORVIEW Light” である．800 $\times$ 600 pixels の画面が実験台上の約 970 $\times$ 720mm の範囲で表示できる．投影光は 500lm の明るさで出力される．

トラッキングビジョン装置

本システムでは富士通製カラートラッキングビジョン”TRV-CU” を使用している．ワークステーションのシリアルポートに接続される．本システムで用いたトラッキングビジョン装置は特定のターゲットをビデオレートで追跡することができる画像処理システムである．相関演算を 33msec の間に 500ヶ所以上で実行できる．カラー画像の R/G/B カラーに対するカラー相関演算を実行でき，目的によってテンプレートサイズの変更，また，対象ブロックの間引きが可能である．全画面 640 $\times$ 480 pixel で画像の取り込みが可能である．

PARTNER において使用するおもな処理方法を以下に示す．

1. 追跡処理

トラッキングビジョン装置のもっとも基本的な機能．事前に画像メモリに登録したテンプレートを登録し，ターゲットが一定の時間にどの方向にどれだけ動いたかを表すことができる．例えば，図 4.12 において，フレームで P^{f-1} にあったターゲットが次のフレームで P^{f} の位置に移動したとすると，P^{f-1} と P^{f} を結ぶ動きベクトルを求め，ターゲットを捕捉することができる．

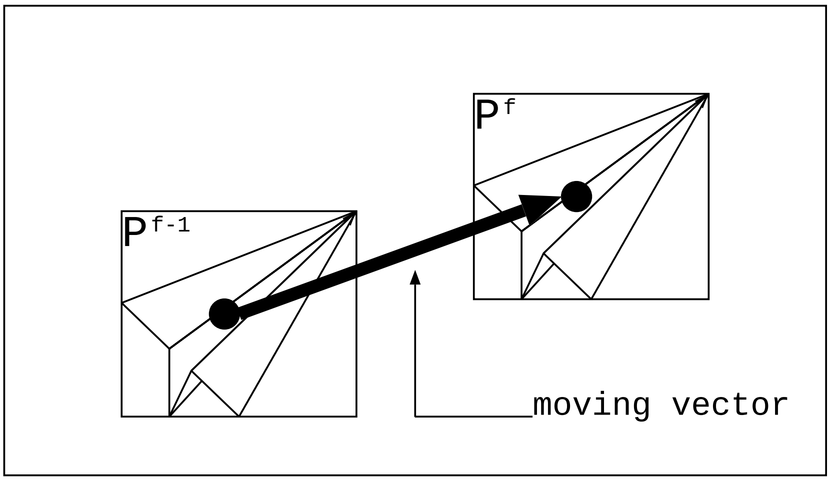


図 4.12: Target tracking

2. テンプレートマッチング

特定の画像パターンをテンプレートとして登録した上で，画面全体を探索すれば，目標パターンが画像内に存在するか，あるいはどこにあるかを判定するパターンマッチングが実行できる．１回の相関演算では $-8 \sim 7$ 画素の範囲しか探索できないが，探索ブロックの位置を順次ずらしながら複数回の相関演算を実行することで，画面全体を探索することができる．図 4.13 では画像メモリに登録した画像をカメラの画面から入力することができる．また，人工的に RGB 情報から画像を生成することも可能である．PARTNER では，部分的に探索エリアを指定し，その中でターゲットを探索できるようにしている．

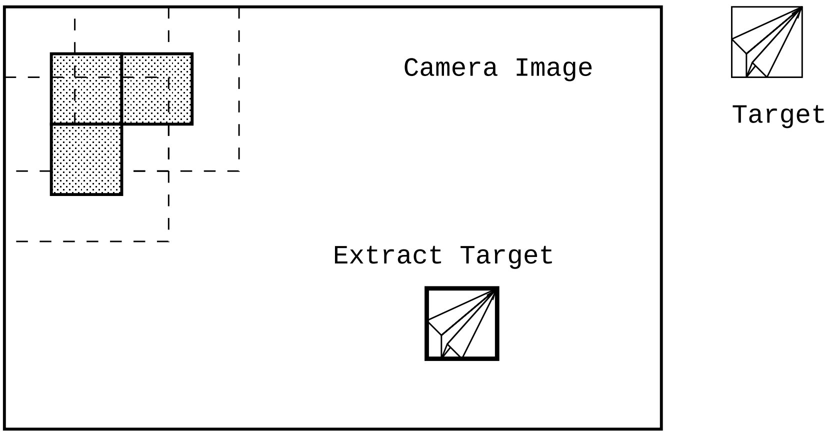


図 4.13: Target tracking

3. 細線化処理

赤外線カメラからのビデオ画像をトラッキングビジョン装置の画像メモリに取り込み，さらにワークステーションに画像情報を取り込む．このデジタル画像データを画像処理パッケージ SPIDER[29] により逐次型アルゴリズムによる 2 値図形の細線化を行なった（図 4.14）．赤外線カメラ画像は 2 値化しやすく，これにより，手の状態認識が有効になる．

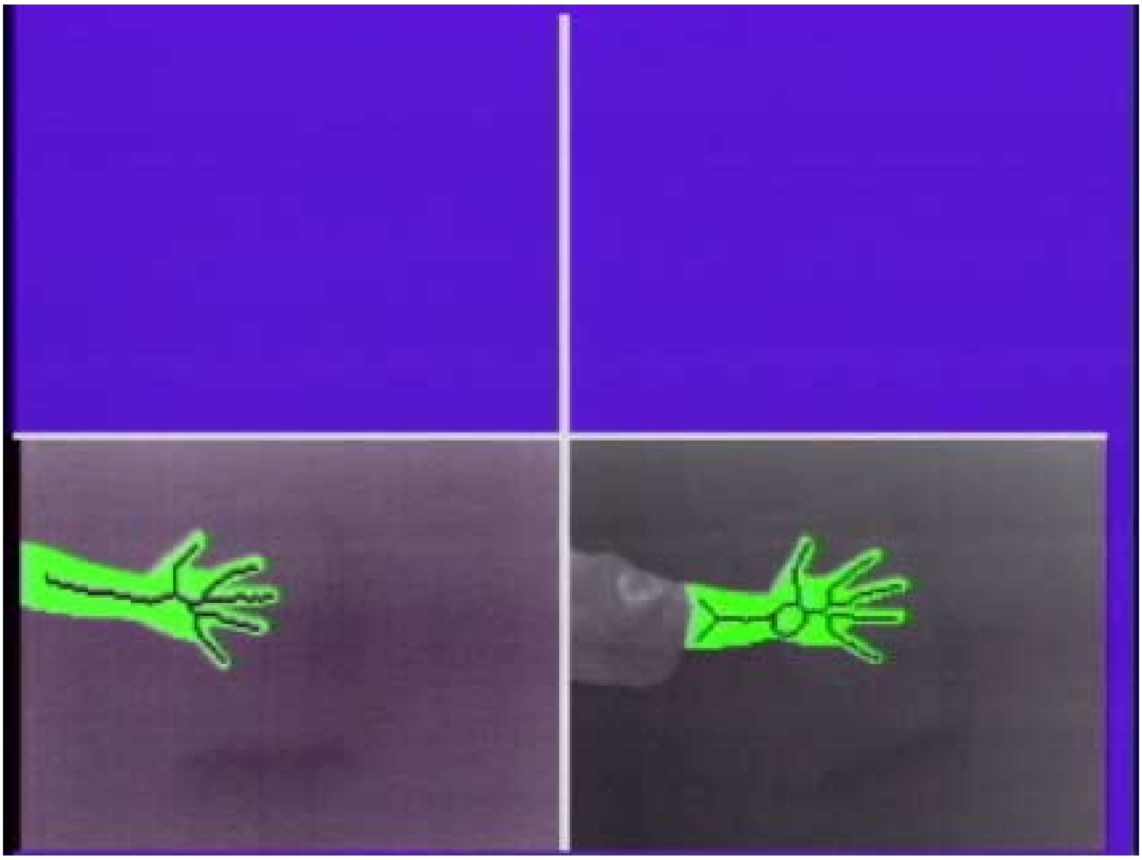


図 4.14: Thinning processing

4.3 組み立てガイド実験

AAF の実験対象として，ポータブル OHP 装置の組み立て作業を選んだ．
組み立て作業は次の手順を必要とする．

(1) head-lock レバーを動かし装置のロックを外す

(2) 本体を持ち上げる

(3) ヘッド部を引き上げ，床に並行になるように回転させる

(4) ミラーを引いて持ち上げ，適当な方向に回転させる

(5) 電源コードのはこのフタを開け，コードを引きだし，プラグにつなぐ

(6) スイッチを入れる



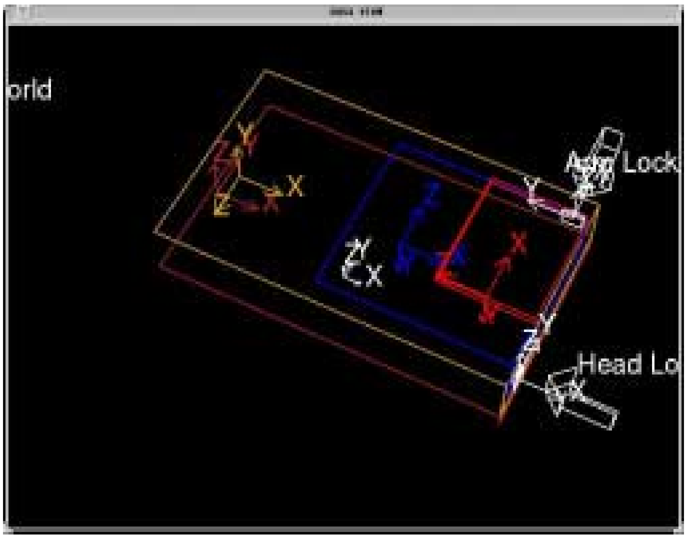
図 4.15: Unfolding of a portable OHP device is a difficult test for a new user without consulting the manual.

(1)-(3) の作業について支援実験を行なった．この OHP はロックが二箇所あり，スタート時点ではどちらを解除して組み立てが始まるのかがわかりづらい．また，ヘッドパーツを引き出し，更に回転させるなどの操作があり，初めての組み立てる人には少々難しい．(図 4.15).

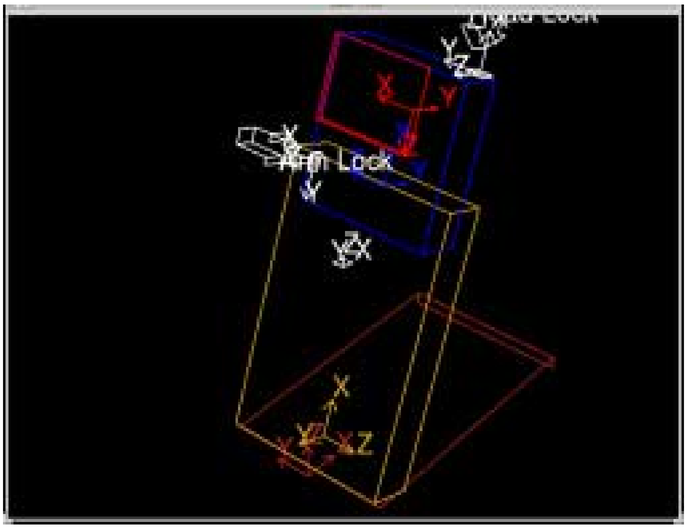
実験を行なうにあたり，事前に OHP の幾何学モデルを作成した．以下に，実験で用いる作業について，シミュレーションで示す (図 4.16).

4.3.1　対象物のモデルマッチング

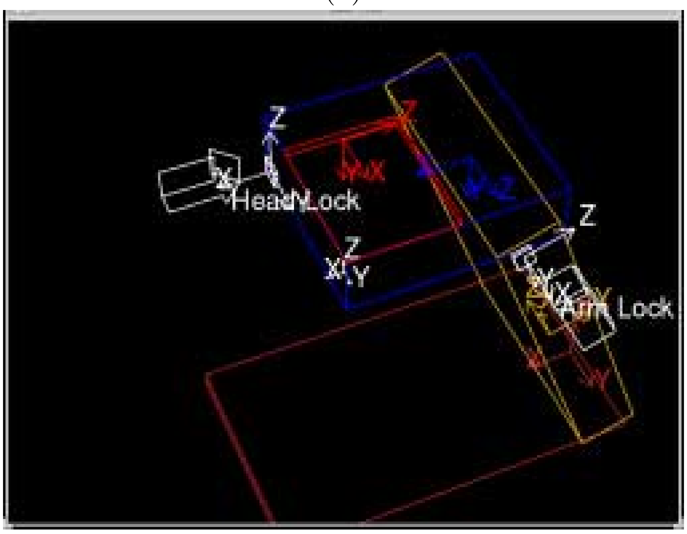
この OHP は平たく，幅が広いため，ランダムドットレンジファインダのみでは，計測が難しい.．モデルの当てはめのために正確に計測することが必要であり，そのためランダムドットステレオレンジファインダとスリット光レンジファインダを協調して働く機能を用いた．具体的には，ランダムドットステレオレンジファインダでラフな形状を抽出する（図 4.17）．その結果に基づいて，図 4.18 はどちらに向いているのかを詳しく調べるために，適切な位置／方向にスリット光を投影し，スリット光レンジファインダにより計測した．中央の縦の線はランダムドットレンジファインダにより得られたおおよその主軸方向をカメラ画像上に描画していて，その線を中点とするようにスリット光を操作して再計測している．小さな四角はスリットの象の検出された場所を示している．これにより，対象物の配置場所を計測し，対象の幾何モデルを実物にマッチングさせる（図 4.19）.



(1)



(2)



(3)

図 4.16: OHP model



図 4.17: A scene of 3D measurement using the random-dot stereo

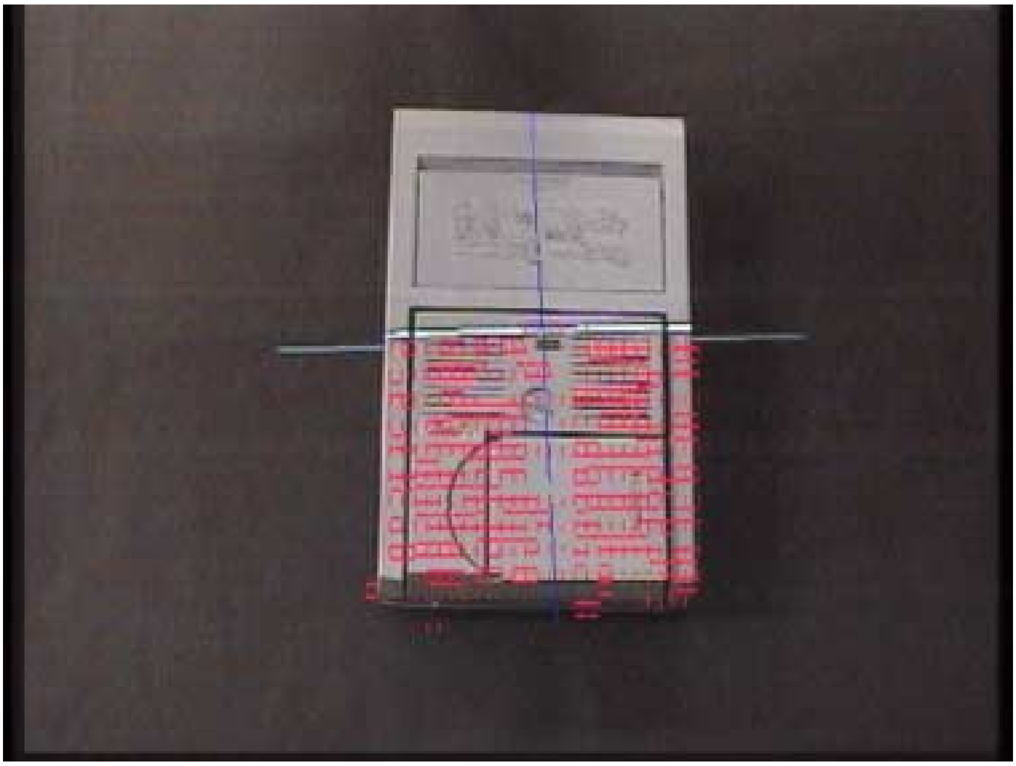


図 4.18: A scene of 3D measurement using light-striped rangefinder (The principal axis is superimposed on the image.)



図 4.19: OHP simulation model is matched by range-finder data

4.3.2 組み立て作業支援の実例

4.3.1 に続き，以下では実際の作業の支援方法について，図 4.1 に従って具体的に述べる．

ガイドオペレーション GO_1

（図 4.20）

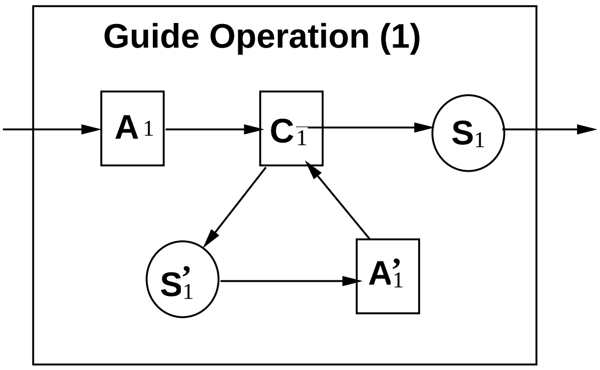


図 4.20: Guide Operation 1 diagram

実験対象である OHP はロックを解除しながら本体を持ち上げる必要がある．しかし，ロックは二ヶ所あるため，どちらが解除すべきロックなのかを表示する必要がある．

1. 実環境において，OHP の幾何モデルを配置する



図 4.21: An infrared image for detecting the human hand. The thinning results and the object model are superimposed.

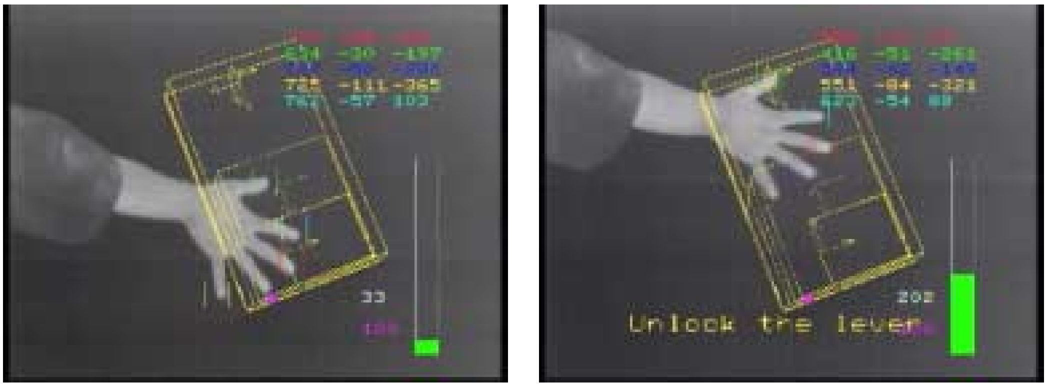


図 4.22: Ranging fingers

2. ロックレバーを配置する
3. 矢印によって位置を示し，ロック解除のためのアノテーション (A_1) を生成する
4. 赤外線画像の中のユーザーの手を探索する
5. 一定時間内にアンロックレバーにその位置が近くなければ (C_1)，ユーザーにアンロック動作を促すアノテーション (A'_1) を生成する

図 4.21 で示すように，手の領域を信頼性良く認識するため，手の細線化処理をする．さらに，指先の位置を検出後，各指に対して追跡を行なう．一定時間経過したら，ロックレバーとの距離を計測する．図 4.22 は指先を赤外線画像により抽出し，OHP のロックの位置（画面中央のピンク色の点）にもっとも近い指との距離をインジケータで示している．もしその距離が満たなければ（図 4.22 右）アンロックレバーを確認していなかったかもしれず，システムが作業修正のためにアノテーションを生成する（図 4.23）．

ガイドオペレーション GO_2

（図 4.24）



図 4.23: Adaptive annotation by monitoring the user's action

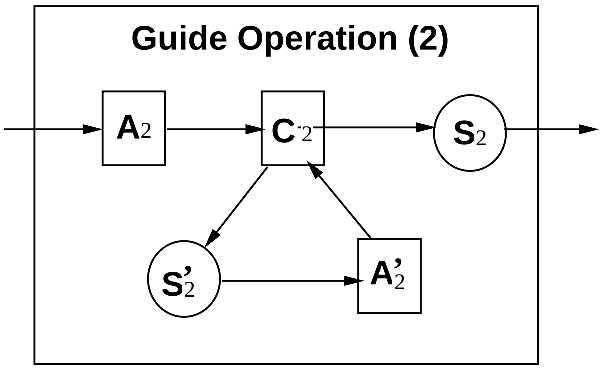


図 4.24: Guide Operation 2 diagram



図 4.25: The message, “Unfold the main body” is projected.

本体のはじめの組み立て作業に関するガイドオペレータ (GO_2) において，次の段階が提供される．

1. アノテーションメッセージ”Unfold the main body”(A_2) を投影する（図 4.25）
2. 本体のエッジの高さをスリット光レンジファインダにより計測する (C_2)（図 4.26）
3. 望ましい状態の幾何学モデルと比較し，本体の高さが十分でなければ（図 4.28(a)），第２のアノテーション”Lift up more”(A_2’) を投影する（図 4.27）
4. 本体の角度が十分である状態 (S_2) が満たされているならば（図 4.28(a)），GO_2 が終了し，GO_3 にシステムは移行する

図 4.28 は持ち上げた高さが適当である時とない時の二つのケースのスリット光レンジファインダから求めた高さデータを示している．対象物の幾何学モデルをシステムは持つので，図の中で示すように操作を点検するため，望ましい状態を予測して計測データとを比較により正確に実行される．

ガイドオペレーション GO_3

（図 4.29）
まず，引き上げの対象であるヘッドパーツを表示し（図 4.32），アノテーションメッセージ”Slide up the head part” を投影する．この操作はヘッドパーツが傾斜角度に沿って引き上げられる．

1. アノテーションメッセージ”Slide up the head part”(A_3) を投影する（図 4.30）．



図 4.26: A striped light is projected to check the height of the main body.

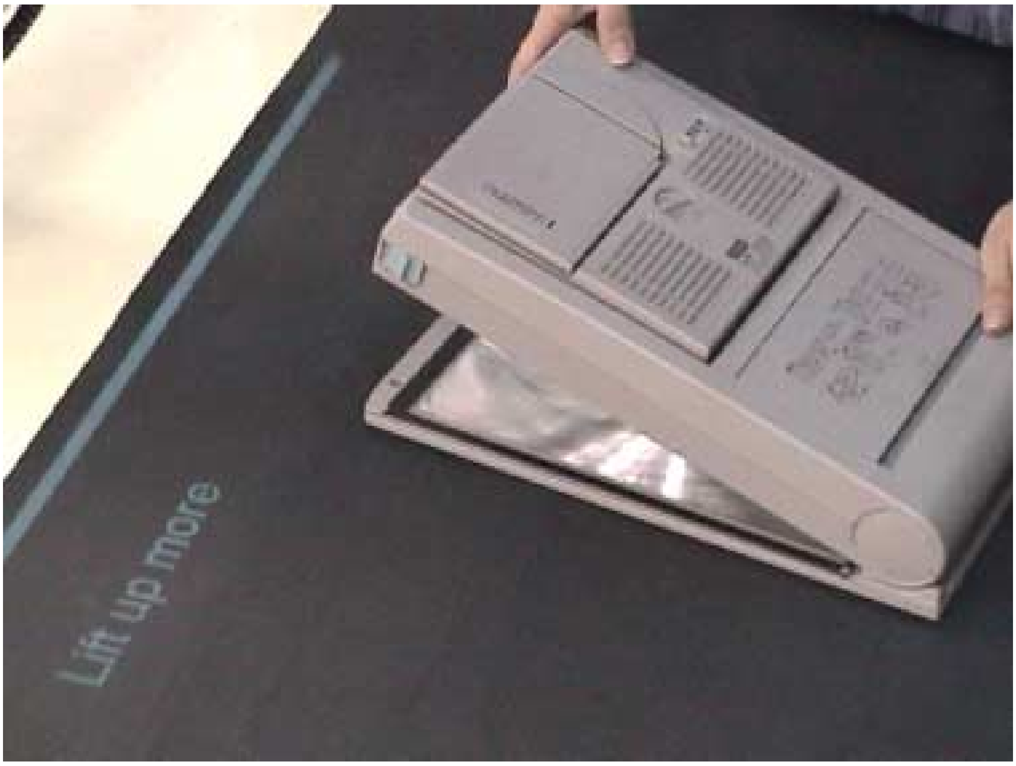


図 4.27: The message, “Lift up more” is projected when the height of the main body is smaller than model.

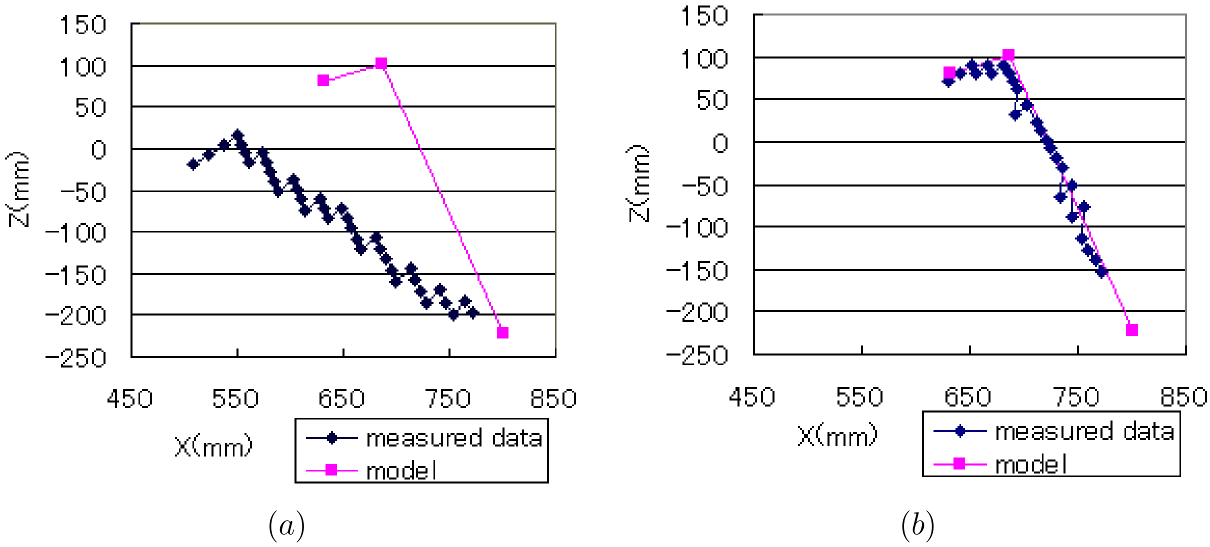


図 4.28: Comparison of the profile of the main body with the model: (a) incorrect, (b) correct operation

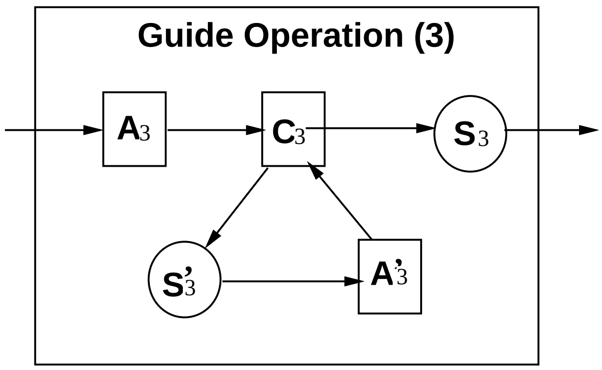


図 4.29: Guide Operation diagram



図 4.30: The message, “Slide up the head part” is projected.

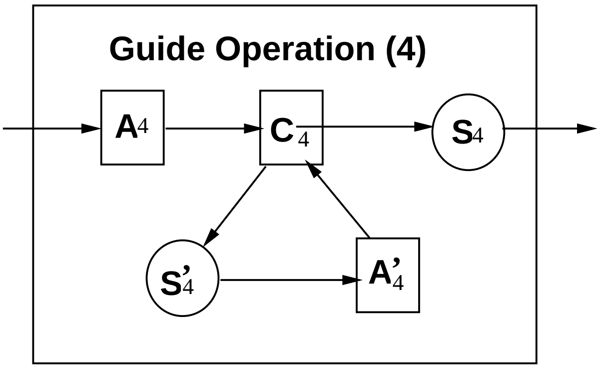


図 4.31: Guide Operation 4 diagram

2. ヘッドパーツの高さをスリット光レンジファインダを使って調べる (C_3).

3. 望ましい状態 (S_3) と望ましくない状態 (S'$_3$) を比較し，望ましくなければ，新しくアノテーションメッセージを投影する (A'$_3$).

ガイドオペレーション GO_4

（図 4.31）
引き上げられたヘッドパーツを並行に回転させるため，初心者には気づきにくい作業である．

1. アノテーションメッセージ”Rotate the head part”(A_4) を投影する．

2. ヘッドパーツの姿勢をスリット光レンジファインダを使って調べる (C_4)（図 4.34）．ヘッドパーツの推測データを図 4.35 に示す．これは望ましくない状態 (a) と望ましい状態 (b) を比較して示している．



図 4.32: The head part is projected to slide up.

3. もし，正しく組み上がっていなければ，第二のアノテーションメッセージ"Rotate the head part to be horizontal"(A'$_4$) を投影する（図 4.33）.

図 4.35 は GO_2 と同様にヘッドパーツの縁の高さから判定し，望ましい状態と望ましくない状態を並べている．

4.3.3　考察

本実験では AAF を用いてポータブル OHP の組み立て実験を行なった．これにより，オペレータへの操作知識の提供が行なうことが確認できた．しかし，人間を作業中に観察する際には手の領域の細線化を用いているが，細線化はすべての指が見え隠れしないことが前提となっている．そのため，この状態では作業をする手の監視を行なうことは難しい．

また，本システムは図 4.1 のアノテーションを行なう表示方法，また，チェックを行なう判定方法などを事前にプログラムに組み込んでいる．これらを自動生成するためには，いくつかの作業の類似性の検出し，それに応じて適当なアノテーションを導き出す必要がある．たとえば，対象物が異なっていても「物体を移動する」といった作業を同じ作業として認識し，その作業に対してオペレータにわかりやすい方法を示すアノテーションを見つける，といった手順を生成する必要がある．これにより，作業支援方法の自動化を計ることができ，対象物，環境に依存しない方法を自動生成を行なうことが期待できる．



図 4.33: The message, "Rotate the head part to be horizontal" is projected.



図 4.34: The state that the head part is properly rotated to be horizontal.

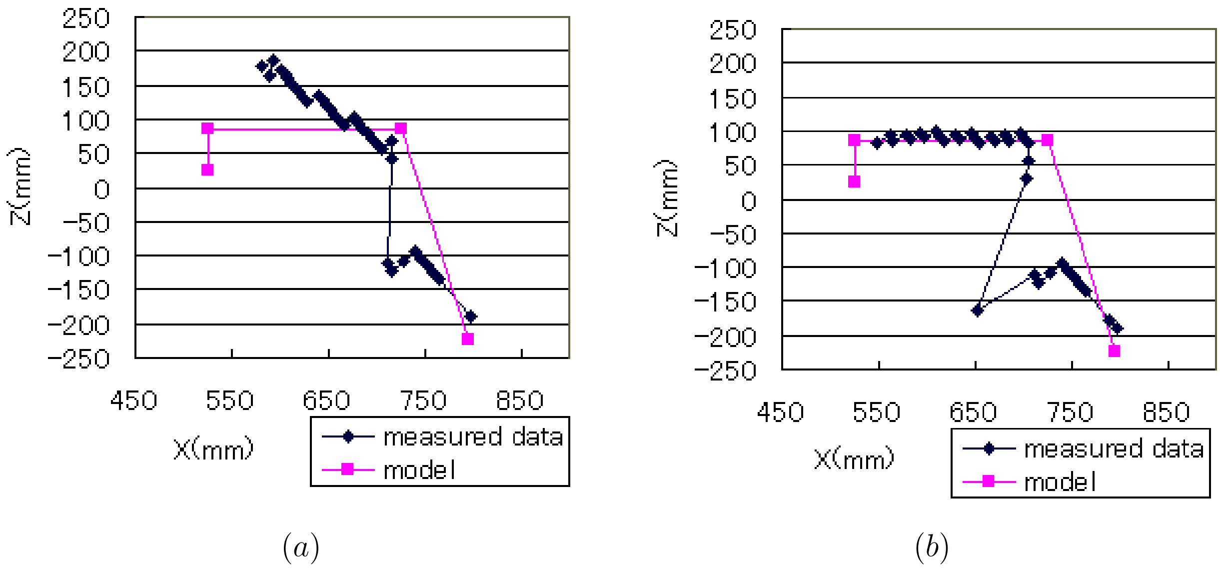


図 4.35: comparison of the profile of the head part with the model : (a) incorrect, (b) correct operation

第5章　結論

柔軟なロボットとの新たなインタラクションシステムの開発を目指し，本論文では，Projector based AR によるヒューマン・ロボットインタフェースシステム:PARTNER の開発の一環として Adaptive Annotation Function を構築した．従来のヒューマン・ロボットインタフェースシステムは主に，人間からロボットへの作業教示を目的に作られたものがほとんどである．しかし，人間へ教えることがあり，その労力を人間に負担させないためにロボットを代役とすることは十分，意義がある．この機能を持たせることにより，人間-ロボット間の柔軟な相互インタラクションシステムの実現が期待できる．

また，人間の作業の支援システムとして AAF は次の特色を持っている．

1. AR 的な枠組みを用いた支援システムを行なうことにより，対象物に直接投影できるため，従来の文書形式のマニュアルに比較して，実物との対応がわかりやすい．
2. 装着型デバイスを用いずに，自然な体勢で作業が行なうことができる．
3. ユーザーと対象オブジェクトの両方を監視し，作業の進行状況に応じて適切なアノテーションが行なうことができる．

本システムをポータブル OHP 機器の組み立て作業に適用し，システムの有効性を確認した．これにより，実際の環境での作業支援の見通しを得ることが出来た．

5.1　今後の課題

本システムでは人の動作を認識するには不十分と思われる．例えば，追跡対象である指先が手の平の中に隠れて見失った場合の処理が行なわれておらず，また，操作対象により手の一部が隠蔽される可能性がある．こうした状況にも対応できる動作監視システムを構築する必要がある．現在，検討している方法は時系列に対して，変化していく手の形を入力データとし，ダイナミックベイジアンネットワーク (DBN) を利用し，動きモデルを推定する方法である．この DBN は確率的に出力されたものとして観測データを扱うため，ノイズにもロバストであり，推測に用いる新たな手法として注目されている．

また，本システムのタスクの支援方法は文字列と簡単な三次元オブジェクトの表示をしているに過ぎず，十分なコミュニケーションが計れているとはいえない．システムの拡充を計るとともに，新たなインタフェースの入出力システムとして音声認識，合成技術の実装を試みている．

以上を行なった上，さらに他のサブシステム，VOP，IIP，IHP との連係を計り，相互インタラクションシステムとしての扱いやすいシステムに汲み上げていく予定である．

付 録A　投影とカメラモデル

本研究は，レンジファインダによって計測された実空間座標系のデータを多く多用する．ここでは，実際の座標系の計測原理を述べるため，対象をカメラ画像面に変換する透視投影とステレオ視による対象の絶対座標値の抽出の仕組みを説明する．

A.1　投影

視覚を情報処理として考える場合は，実際のカメラの物理的な投影を考える必要はなく，カメラの幾何学的な性質のみを考えれば良い．実際のカメラはレンズを用いて，撮像素子に集光させているに過ぎない (図 A.1)．そのため，これから話す内容は光学素子に至るすべての光がレンズ中央部である**光学中心**(optical center) のみを通過してくるカメラ，**ピンホールカメラ**(pinhole camera) をベースに考える．

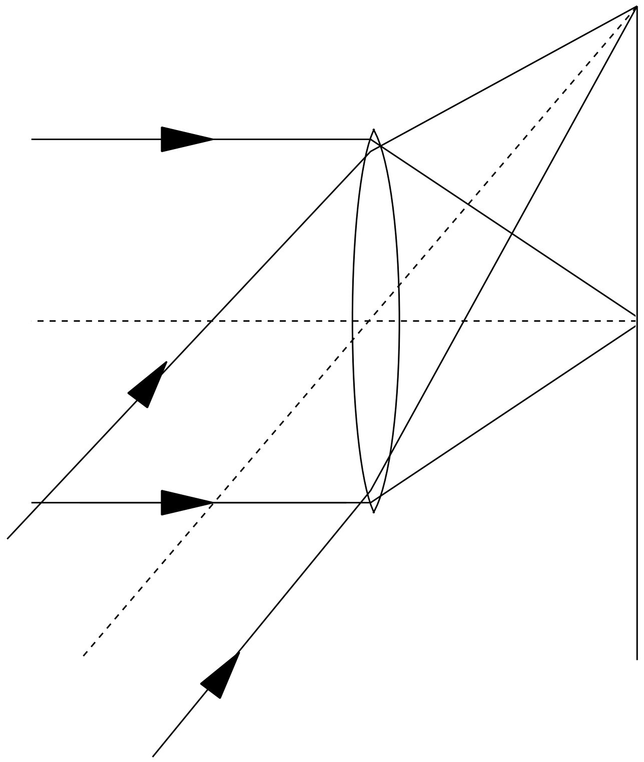


図 A.1: projection using a lens

A.1.1 透視投影

光学中心 C から**画像面**に垂直に引いた直線 l_0 を光軸，光軸と画像面との交点 $\boldsymbol{x}_c$ を画像中心と呼ぶ．図 A.2 に示すように，ピンホール・カメラの光学中心を座標原点にとり，３次元座標の X 軸と Y 軸をそれぞれ画像面上の x 軸および y 軸に平行にとり，Z 軸を光軸 l_0 に対し，平行にとることにする．このようにカメラを中心として定義した座標系のことを**カメラ座標系**(camera coordinate system) と呼ぶ．

いま，３次元空間中での座標点 $\boldsymbol{X} = [X, Y, Z]^\top$ が 2 次元画像上の点 $\boldsymbol{x} = [x, y]^\top$ に投影されているとする．この画像上の点を３次元空間中の点と考えるとその位置は $(x, y, 1)$ である．図 A.2 よりピンホール・カメラによる投影では，画像上の点 x と $f(= 1)$ の比が X と Z の比に等しく，また，y と Z の比が等しい．したがって，３次元空間中の点とその投影像との間には以下の関係が成り立つ．

$$x = \frac{X}{Z}, y = \frac{Y}{Z} \tag{A.1}$$

このような投影を透視投影 (perspective projection) と呼ぶ．式（A.1）より透視投影は Z に関し非線形であることがわかる．

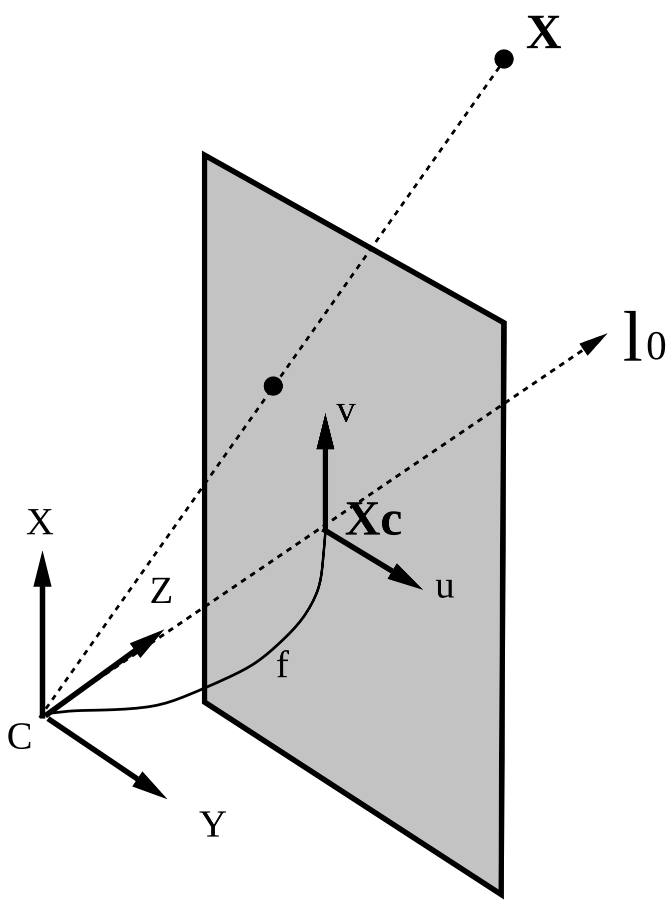


図 A.2: Camera coordinate system

A.1.2　斉次座標

透視投影の大きな特徴は平行な複数の投影像が画像面上で一つの点で交わることである．ユークリッド幾何では，平行な直線同士はこれらを無限に伸ばしても決して交差することはないとしている．しかし，これらを透視投影すると投影像においてこれらの点 (**消失点**:vanishing point) は一つになってしまい，空間と投影像との間での一体一対応が崩れてしまう．この問題を避けるため，透視投影を扱う**射影幾何**(projective geometry) では平行線は実は無限遠で交わっていると考える．



図 A.3: vanishing point

射影空間では無限遠要素を持つため，ユークリッド空間とは異なる構造を持つ必要がある．たとえば，ユークリッド直線上の点の位置は一つの実数 x であらわせる．しかし，ここではより広い空間を表すため，このような点 x を二つの比によって表すことにする．すなわち，

$$x = \frac{x_1}{x_2} \tag{A.2}$$

を満たすような二つの実数の組 $\tilde{x}$ を考える．

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{A.3}$$

ただし，x_1 および x_2 のうちどちらか一方は必ず 0 でないとする．このとき，もし，$x_2 \neq 0$ であれば，式 A.2 より x は常に有限な実数となるから，$\tilde{x}$ が通常のユークリッド直線上の点を表すことがわかる．一方，もしも，x_2 である場合には，式 A.2 の x は無限大となる．すなわち，この場合には，$\tilde{x}$ は直線上の無限遠点を表す.

このように空間の次元を一つ上げることにより，射影直線上のすべての点を表現できる．このように空間の次元を一つ上げ，これらの比によって空間を表現するような座標の取り方を**斉次座標**(homogeneous coordinates) あるいは，**同次座標**と呼ぶ．以降，斉次座標には˜をつけて表すことにする.

ここで，$\tilde{x}$ を 0 以外の実数である λ 倍した

$$\lambda\tilde{x} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} \tag{A.4}$$

は，以下の式が成り立つ.

$$\frac{\lambda x_1}{\lambda x_2} = \frac{x_1}{x_2} = x \tag{A.5}$$

つまり，$\lambda\tilde{x}$ は $\tilde{x}$ と同一であり，このような $\tilde{x}$ と $\lambda\tilde{x}$ との関係を**同値関係**(equibalence relation) と呼び，

$$\tilde{x} \sim \lambda\tilde{x} \tag{A.6}$$

で表す.

A.2 透視カメラモデル

実際のカメラを透視投影によってモデル化したものを**透視カメラモデル**と呼ぶ．ここでは斉次座標を使うことにより，非線形の透視投影を行列を使って考える.

A.2.1 斉次座標による透視投影

画像面を 2 次元射影平面とみなし，画像面上の点 $\boldsymbol{x} = [x, y]^\top$ に対して，同値関係

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \sim \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{A.7}$$

が成り立つような斉次座標 $\tilde{\boldsymbol{x}} = [x_1, x_2, x_3]^\top$ をとることにする．同様に 3 次元空間は射影空間とみなし，3 次元点 $\boldsymbol{X} = [X, Y, Z]^\top$ に対して次に示す同値関係が成り立つような成り立つような斉次座標 $\tilde{\boldsymbol{X}} = [X_1, X_2, X_3, X_4]^\top$ を割り当てる.

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \sim \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{A.8}$$

このとき，式 (A.1) の透視投影は斉次座標 $\tilde{\boldsymbol{x}}$，$\tilde{\boldsymbol{X}}$ を使うと次のような行列演算の形で表すことができる．

$$\lambda \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \tag{A.9}$$

ここで，λ は任意の実数である．今後の式の展開のために式 (A.9) を次のように記号を使って表す．

$$\lambda \tilde{\boldsymbol{x}} = \boldsymbol{P}_f \tilde{\boldsymbol{X}} \tag{A.10}$$

ここで，

$$\boldsymbol{P}_f = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{A.11}$$

である．

A.3 物理的な座標と画像座標

CCD カメラで得られる画像はディジタル化され，コンピュータの内部で**画素**(pixel) と呼ばれる単位で記録される．一つの画素が一つの明るさを持ち，これが縦横にメッシュ上に並んだものが画像データである．この画像データの原点をどこにとるか，画素のメッシュの区切り方 (メッシュの細かさや縦横比) をどうするかなどは撮像系の設計次第で変更できる．カメラから得る画像データは，このような物理的な大きさや位置とは無関係な画素単位の画像座標に基づいている．したがって，実際のカメラをモデル化する場合には，物理的座標 $\boldsymbol{x}$ から画像座標 $\boldsymbol{m}$ への変換を考えなければならない．このような変換は原点位置合わせのための平行移動，縦横のスケール変換および焦点距離に応じたスケール変換を組み合わせたものであり，次のように記述できる．

$$\tilde{\boldsymbol{m}} = \boldsymbol{A} \tilde{\boldsymbol{x}} \tag{A.12}$$

ここで，$\tilde{\boldsymbol{m}} = [m_1, m_2, m_3]^\top$ は画像座標 $\tilde{\boldsymbol{m}} = [u, v]^\top$ の斉次座標であり，

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \sim \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{A.13}$$

なる同値関係が成り立つようにとるものとする．すなわち，$u = m_1/m_3, v = m_2/m_3$ である．また，$\boldsymbol{A}$ は物理的座標 $\boldsymbol{x}$ を画像座標 $\boldsymbol{m}$ へ変換する行列であり，焦点距離 f，画像中心の座標 (u_0, v_0)，u および v 方向のスケール・ファクター k_u，k_v およびせん断係数 k_s によって次のように表される．

$$\boldsymbol{A} = \begin{bmatrix} fk_u & fk_s & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.14}$$

式 (A.10) および，式 A.12 より 3 次元の点 $\boldsymbol{X}$ は画素 $\boldsymbol{m}$ に以下のように投影される．

$$\lambda\tilde{\boldsymbol{m}} = \boldsymbol{A}\boldsymbol{P}_f\tilde{\boldsymbol{A}} \tag{A.15}$$

このとき，$\boldsymbol{A}$ は**カメラ校正行列**(camera calibration matrix) と呼ばれる．これは式 (A.14) から明らかなように，カメラの**内部パラメータ**(intrinsic parameters matrix) によってとも呼ばれる．$\boldsymbol{A}$ のなかでは f と k_u，k_v，k_s とは掛け合わされた形となっているため，カメラを外側から見ると f と k_u，f と k_v をそれぞれ一つの変数とみなし，$\boldsymbol{A}$ を次のような自由度 5 の上三角行列 (upper triangular matrix) により表すことにする．

$$A = \begin{bmatrix} a_u & s & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.16}$$

ここで，$a_u = fk_u$，$a_v = fk_v$，$s = fk_s$ である．このような $\boldsymbol{A}$ は直角性が保たれないような画像の歪みも扱うことができる．

A.4　カメラ座標とワールド座標

前節はカメラを中心としたカメラ座標を使ったが，今度は図 A.4 に示すような複数カメラや対象部つかんとの関係を考える．すべてのカメラや対象物に対して共通に決められた基準の座標のことを**ワールド座標**(world coordinates) と呼ぶ.

ワールド座標 $\boldsymbol{X}_w$ からカメラ座標 $\boldsymbol{X}$ へは 3 次元の回転 $\boldsymbol{R}$ および並進 $\boldsymbol{T}$ によって次のように変換される．

$$\boldsymbol{X} = \boldsymbol{R}\boldsymbol{X}_w + \boldsymbol{T} \tag{A.17}$$

3 次元の回転や並進はそれぞれ次のように X 軸回り，y 軸回り，z 軸回りの回転 θ_X，θ_Y，θ_Z および X 方向，Y 方向，Z 方向の並進 T_X，T_Y，T_Z よりなる．

$$\begin{aligned} \boldsymbol{R} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_X & -\sin\theta_X \\ 0 & \sin\theta_X & \cos\theta_X \end{bmatrix} \begin{bmatrix} \cos\theta_Y & 0 & \sin\theta_Y \\ 0 & 1 & 0 \\ -\sin\theta_Y & 0 & \cos\theta_Y \end{bmatrix} \begin{bmatrix} \cos\theta_X & -\sin\theta_X & 0 \\ \sin\theta_X & \cos\theta_X & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \boldsymbol{T} &= \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} \end{aligned} \tag{A.18}$$

式 A.17 を斉次座標を使えば次のように表される．

$$\tilde{X} = \boldsymbol{M}\tilde{X}_w \tag{A.19}$$

ここで，M は回転 R と並進 T を組み合わせた 4×4 行列であり，次の通りである．

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0^T & 1 \end{bmatrix} \tag{A.20}$$

$\boldsymbol{M}$ はワールド座標に対するカメラの姿勢 (回転) および位置 (並進) によって決まる行列である．回転や並進はカメラの**外部パラメータ**(extrinstic parameter) と呼ばれ，内部パラメータ行列 $\boldsymbol{A}$ に対して $\boldsymbol{M}$ のことをカメラの**外部パラメータ行列**(extrinsic parameters matrix) と呼ぶ.

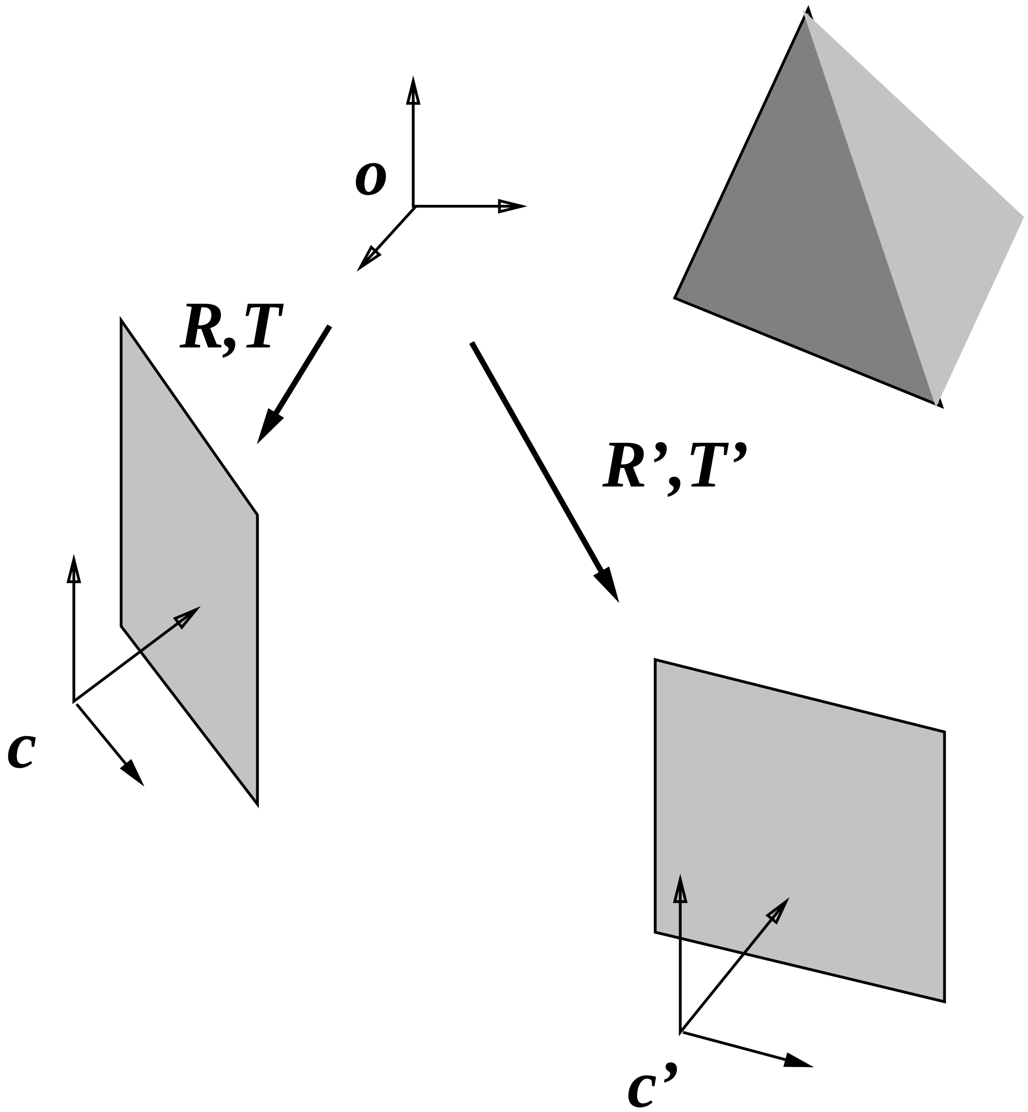


図 A.4: Camera coordinates and world coordinate

A.5 透視カメラモデルと射影カメラモデル

式 (A.15) 及び (A.19) より，ワールド座標系における 3 次元空間の点 $\tilde{\boldsymbol{X}}$ は 2 次元画像上の点 $\tilde{\boldsymbol{m}}$ につぎのように投影されることがわかる．

$$\lambda\tilde{\boldsymbol{m}} = \boldsymbol{P}\tilde{\boldsymbol{X}} \tag{A.21}$$

ここで $\boldsymbol{P}$ は

$$\boldsymbol{P} = \boldsymbol{A}\boldsymbol{P}_f\boldsymbol{M} = \boldsymbol{A}[\boldsymbol{R} \quad \boldsymbol{T}] \tag{A.22}$$

であらわされる．

式 (A.21) で表されるカメラモデルを**透視カメラモデル**(perspective camera model) と呼ぶ．$\boldsymbol{P}$ は**透視カメラ行列**(perspective camera matrix) と呼ばれ，カメラの内部パラメータおよび外部パラメータにより構成される．

また，さらに一般化し，カメラ行列 $\boldsymbol{P}$ を一般の 3×4 行列で表すことにする．

$$\lambda \tilde{\boldsymbol{m}} = \boldsymbol{P}_P \tilde{\boldsymbol{X}} \tag{A.23}$$

ここで，$\boldsymbol{P}_p$ は次のような 3×4 行列である．

$$\boldsymbol{P}_P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \tag{A.24}$$

このように一般の 3×4 行列で表されたカメラモデルのことを**射影カメラモデル**と呼び，$\boldsymbol{P}_P$ のことを**射影カメラ行列**と呼ぶ．

A.6 校正済みカメラによる射影変換と３次元座標の推定

空間中の点と画像上の点の斉次座標をそれぞれ $[X_1, X_2, X_3, X_4]^T$=$[X, Y, Z, 1]^T$,$[m_1, m_2, m_3]^T$=$[u, v, 1]^T$ と置いて，次のような射影カメラを考える．

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{A.25}$$

内部パラメータ，外部パラメータが直接わからなくても，カメラ行列が既知であれば，この式より，空間中の点 $\boldsymbol{X} = (X, Y, Z)$ のカメラの投影像 (u, v) が計算できる．

また，A.25 より，λ を消去し，X，Y，Z に関してまとめると次のようになる．

$$\begin{bmatrix} p_{31}u - p_{11} & p_{32}u - p_{12} & p_{33}u - p_{13} \\ p_{31}v - p_{21} & p_{32}v - p_{22} & p_{33}v - p_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{14} - p_{32}u \\ p_{24} - p_{34}v \end{bmatrix} \tag{A.26}$$

これは一つの視点の画像からの拘束式であり，ステレオカメラで二つの視点から計測すると以下の拘束式が得られる．

$$\begin{bmatrix} p_{31}u - p_{11} & p_{32}u - p12 & p_{33}u - p_{13} \\ p_{31}v - p_{11} & p_{32}v - p22 & p_{33}v - p_{23} \\ p'_{31}u' - p'_{11} & p'_{32}u' - p'_{12} & p'_{33}u' - p'_{13} \\ p'_{31}v' - p'_{21} & p'_{32}v' - p'_{22} & p'_{33}v' - p'_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{14} - p_{32}u \\ p_{24} - p_{34}v \\ p'_{14} - p'_{32}u' \\ p'_{24} - p'_{34}v' \end{bmatrix} \tag{A.27}$$

この式は未知数３に対して拘束が四つあるので，これを

$$MX = b \tag{A.28}$$

とおくと，X の最小二乗解は M の一般化逆行列 M^+ を使って次のように求まる．

$$X = M^+ b \tag{A.29}$$

これにより，カメラ行列が既知の二つのカメラに変換される対象の投影像 (u, v)，(u', v') の座標から対象のワールド座標値が算出できる．

付 録B　スリット光レンジファインダによる絶対座標点抽出

本システムに実装されているスリット光レンジファインダは，プロジェクタの画像面の情報とCCDカメラの検出座標によりワールド座標値を計測する．本章では，そのワールド座標の計算方法について述べる．

B.1　画像面上の直線の斉次座標表現

ユークリッド幾何において，２次元平面上で (u_1, v_1)，(u_2, v_2) を結ぶ線分を考える．あらかじめ，a，b，c に次の値を与える．

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} u_2 - u_1 \\ v_1 - v_2 \\ u_2 v_1 - u_1 v_2 \end{pmatrix} \tag{B.1}$$

ここで，直線上の点 (u', v') は

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} + t \begin{pmatrix} a \\ b \end{pmatrix} \tag{B.2}$$

となる．

ワールド座標上の点 (X, Y, Z) が画像面に射影変換された時，上記の線分を通るとすると以下の式は画像面上の直線への投影を示している．

$$\lambda \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \boldsymbol{P} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{B.3}$$

ただし，λ は任意の実数，$\boldsymbol{P}$ は画像面の射影カメラ行列を示し，

$$\boldsymbol{P} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \end{pmatrix}$$

である．

式 (B.2)(B.3) から，u'，v'，t，を消去すると次の式が成り立つ．

$$\lambda \begin{pmatrix} c \\ 1 \end{pmatrix} = \begin{pmatrix} p_{00}b + p_{10}a & p_{01}b + p_{11}a & p_{02}b + p_{12}a & p_{03}b + p_{13}a \\ p_{20} & p_{21} & p_{22} & p_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{B.4}$$

B.2　絶対座標点抽出方法

前節で導いた画像面上の直線をスリット光のデータとして距離計測に利用する．プロジェクタ上で線分 $(u1, v1)$，$(u2, v2)$ を描画し，投影された線分上の点を $w = (x, y, z)$ を走査する．このとき，その点が CCD カメラの画像面上で (U, V) で検出されたとする．（図 B.1）

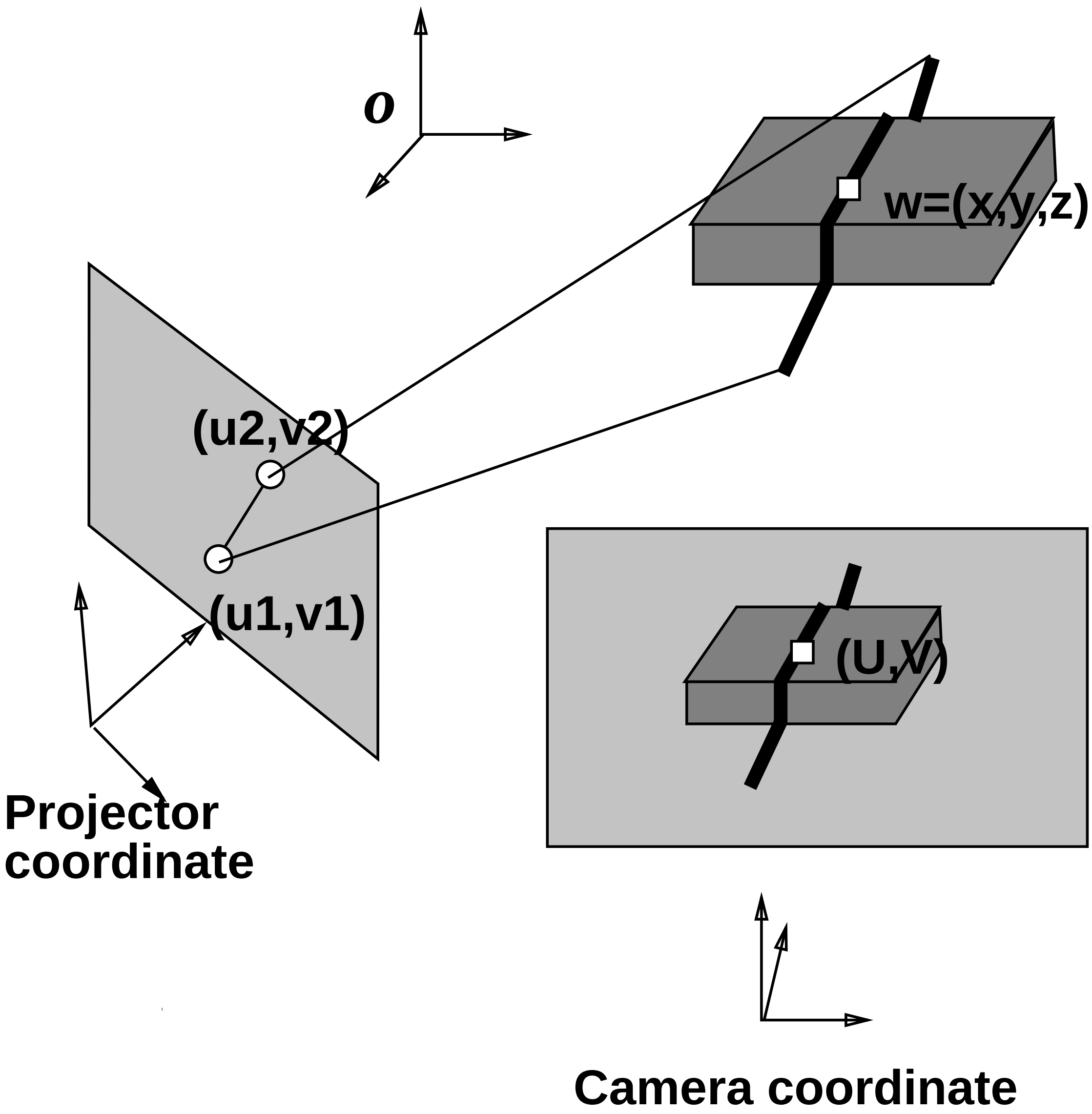


図 B.1: Striped-light range-finder coordinates system

このとき，CCD カメラの画像面では以下の式が成立する．

$$\lambda' \begin{pmatrix} U \\ V \\ 1 \end{pmatrix} = \boldsymbol{C} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{B.5}$$

ここで，$\boldsymbol{C}$はカメラモデルの射影カメラ行列

$$\boldsymbol{C} = \begin{bmatrix} c_{00} & c_{01} & c_{02} & c_{03} \\ c_{10} & c_{11} & c_{12} & c_{13} \\ c_{20} & c_{21} & c_{22} & c_{23} \end{bmatrix}$$

であり，λ' は任意の実数を示す．この式と，式（B.4）からスリット像の投影され，CCD カメラで検出された点は以下の式を満たす.

$$\begin{pmatrix} \lambda c \\ \lambda \\ \lambda' U - c_{03} \\ \lambda' V - c_{13} \\ \lambda' \end{pmatrix} = \begin{pmatrix} p_{00}b + p_{10}a & p_{01}b + p_{11}a & p_{02}b + p_{12}a & p_{03}b + P_{13}a \\ p_{20} & p_{21} & p_{22} & p_{23} \\ c_{00} & c_{01} & c_{02} & c_{03} \\ c_{10} & c_{11} & c_{12} & c_{13} \\ c_{20} & c_{21} & c_{22} & c_{23} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{B.6}$$

ここで，λ, λ' を消去し，(x, y, z) に関してまとめると以下の式になる.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p_{00}b + p_{10}a - p_{20}c & p_{01}b + p_{11}a - p_{21}c & p_{02}b + p_{12}a - p_{22}c \\ c_{20}U - c_{00} & c_{21}U - c_{01} & c_{22}U - c_{02} \\ c_{20}V - c_{10} & c_{21}V - c_{11} & c_{22}V - c_{12} \end{pmatrix}^{-1} \begin{pmatrix} p_{23}c - p_{03}b - p_{13}a \\ c_{03} - p_{23}U \\ c_{13} - p_{23}V \end{pmatrix} \tag{B.7}$$

この式から，プロジェクタの直線情報と CCD カメラでとらえた座標値からワールド座標値 (x, y, z) が計測できる.

関連図書

[1] Virtual reality special report, 1996. http://www.boeing.com/assocproducts/art/tech_focus.html.

[2] Viavoice 98 日本語版. 日本 IBM 株式会社, 1998.

[3] Bajura, Fuchs, and Ohbuchi. Merging virtual reality with the real world:seeing ultrasound imagery within the patient. In *Compter Graphics*, Vol. 26, pp. 203–210. Proceedings of SIGGRAPH '92, 1992.

[4] D.Drascic and P.Milgra. Stereoscopic vision and augmented reality. *Scientific Computing and Automation*, Vol. 9, No. 7, pp. 31–34, June 1993.

[5] Feiner, Macltyre, and Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, Vol. 36, No. 7, pp. 53–62, July 1993.

[6] K.Ogawara, S.Iba, T.Tanuki, H.Kimura, and K.Ikeuchi. Recognition of human task by attention point analysis. Vol. 3, pp. 2121–2126, Kagawa, November 2000. International Conference on Intelligent Robot and Systems (IROS)'00.

[7] M.Nagao and T.Matsuyama. Edge preserving smoothing. In *CGIP*, Vol. 9, pp. 394–407, 1979.

[8] Vladimir Ivan Pavlovic. *Dynamic Bayesian Networks for Information Fusion with Application to Human-Computer Interfaces.* PhD thesis, University of Illinois at Urbana-Champaign, 1999.

[9] F. K. H. Quek. Toward a vision-based hand gesture interface. pp. 17–31. IEEE Int. Conf. Virtual Reality Software and Technology Conf, 1994.

[10] F. K. H. Quek. Eyes in the interface. *Image and Vision Computing*, Vol. 13, pp. 78–91, Augest 1995.

[11] R.T.Azuma. A survey of augmented reality. *PRESENCE*, Vol. 6, No. 4, pp. 355–385, 1997.

[12] S.Sato and S.Sakane. A human-robot interface using an interactive hand pointer that projects a mark in the real work space. pp. 589–595, San Francisco, California, April 24-28 2000. IEEE International Conference on Robotics and Automation.

[13] Thad Starner and Alex Pntland. Real-time american sign language recognition from video using hidden markov models. Technical Report 375, M.I.T Media Laboratory Perceptual Computing Section, 1995.

[14] 寺島, 坂根. 拡張ディジタルデスクを用いるヒューマン・ロボットインタフェース. 日本ロボット学会誌, Vol. 16, No. 8, pp. 1091–1098, 1998.

[15] T.Sato, E.Inoue, and H.Mizoguchi. Support in situ for operation of fabrication. pp. 535–538. ROBOMEC'95, 1995.

[16] Pierre Wellner. Interactive with paper on the digitaldesk. *Communications of the ACM*, Vol. 36, No. 7, pp. 87–97, July 1993.

[17] Y.Ohta and H.Tamura. *Mixed Reality -merging real and virtual worlds-*. Springer, 1999.

[18] Pierre Wellner, et al. 電能強化環境. パーソナルメディア, 東京都品川区平塚 1-7-7 MY ビル, 1995. 坂村健 監訳.

[19] 小林, 佐藤, 小池. Enhanced desk のための赤外線画像を用いた実時間指先認識インターフェース. In *Workshop on Interactive Systems and Software (WISS'99)*, pp. 49–54. 日本ソフトウェア科学会, 近代科学社, December 1999.

[20] 小林, 小池, 佐藤. 赤外線画像を用いた指先実時間追跡による enhanced desk の実現. ヒューマンインターフェース学会　ヒューマンインターフェースシンポジウム HIS9, pp. 417–422. ヒューマンインターフェース学会, Octobe 1999.

[21] 佐藤信. 実空間にマークを投影するインタラクティブハンドポインタの構成. Master's thesis, 中央大学大学院, 理工学研究科経営システム工学専攻, 2000.

[22] 佐藤淳. コンピュータビジョン -視覚の幾何学-, 第 3.1 章, pp. 48–49. コロナ社, 1999.

[23] 松井俊浩. オブジェクト指向型モデルに基づくロボットプログラミングシステムの研究, 電子技術総合研究所研究報告, 第 926 巻, 第 4 章, pp. 87–125. April 1991.

[24] 橋本ほか. ランダムドットパターン投光ステレオを用いた物流ロボット視覚システムの開発. 日本ロボット学会誌, Vol. 17, No. 1, pp. 48–49, 1999.

[25] 伴好弘. 強調現実感による工業作業支援システム. PhD thesis, 奈良先端科学技術大学大学院, 情報科学研究科情報処理学専攻, May 24 1999.

[26] 荒木, 田中, 藤野, 佐藤. 高速三次元形状計測装置の試作. 電子情報通信学会論文誌, Vol. J71-D, No. 10, pp. 2059–2068, October 1988.

[27] 山下, 佐藤, 坂根. 実空間強化とヒューマン・ロボットインタフェースの研究-対話型レンジファインダ機能の構成-. システム・制御研究会資料 SC00-4, pp. 19–22. 電気学会, May 2000.

[28] 小関, 中野, 山本. 光切断法による距離検出の実時間化. 電子情報通信学会論文誌, Vol. J68-D, No. 5, pp. 1141–1148, 1985 1985.

[29] 情報処理振興事業協会. 画像処理サブルーチン・パッケージ SPIDER-II. 情報処理振興事業協会 協同システム開発株式会社, 1986.

[30] 森田, 沢崎, 内山, 佐藤. カラートラッキングビジョン. 第１４回日本ロボット学会学術講演会, pp. 279–280. 日本ロボット学会, 1996.

[31] 西村広光, 小林誠, 丸山稔, 中野康明. 多方向特徴抽出 hmm と bagging による多数決を利用した文字認識. 電子情報通信学会論文誌, Vol. J82-DII, pp. 1429–1434, 1999.

[32] 中西, 山岸, 梅田. 画像処理によるインテリジェントルーム支援システム. 第１８回日本ロボット学会学術講演会予稿集, 第 1 巻, pp. 397–398, September 2000.